DOCUMENT RESUME

ED 314 789 CS 507 008

AUTHOR Arter, Judith A.
TITLE Assessing Communication Competence in Speaking and Listening: A Consumer's Guide.
INSTITUTION Northwest Regional Educational Lab., Portland, Oreg.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Nov 89
CONTRACT 400-86-0006
NOTE 129p.; Photoreduced type in Appendix E may not reproduce well.
AVAILABLE FROM Test Center, Northwest Regional Education Laboratory, 101 S.W. Main Street, Suite 500, Portland, OR 97204.
PUB TYPE Guides - Non-Classroom Use (055) -- Reference Materials - Directories/Catalogs (132)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Communication Skills; *Evaluation Criteria; *Interpersonal Competence; Language Arts; *Listening; Listening Skills; *Speech Communication; Speech Skills; *Student Evaluation; Testing

ABSTRACT

Many educational institutions are currently attempting to improve their attention to speaking and listening in curriculum and instruction, and this guide is intended to assist educators to become more knowledgeable about tools for assessing speaking and listening. Designed for those somewhat knowledgeable in the areas of assessment and language arts instruction, the guide has three major sections. The first section, a short discussion of assessing speaking and listening, includes information about definitions, taxonomies of skills, issues in validity and reliability, the current state-of-the-art in assessing speaking and listening, and what to consider when selecting an assessment tool. The second section contains descriptions and reviews of assessment tools, providing longer reviews for instruments that are readily accessible, that measure some aspect of "communication competence," and that have some technical information available. These longer reviews are evaluative and include descriptions of the purpose(s) the author sees for the assessment instrument; content (tasks, response modes, and scoring procedures); reliability; validity studies; amount of help with interpretation and use; and source. The second section also presents short reviews of research instruments, achievement test series, instruments developed by educational agencies and instruments lacking technical investigation. The final section lists additional resources available to the user such as print resources and professional organizations. The guide also includes a summary table of all instruments reviewed, a 64-item bibliography, a glossary, and an index so that instruments can be easily located. (SR)

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: CS

In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view. TM

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Jerry D. Kirkpatrick

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

November 1989

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement (OERI), U.S. Department of Education under contract Number 400-86-0006. The content of this publication does not necessarily reflect the views of the department or any agency of the U.S. government.

ASSESSING COMMUNICATION COMPETENCE IN SPEAKING AND LISTENING

A CONSUMER'S GUIDE

Judith A. Arter

Test Center
Northwest Regional Educational Laboratory
101 S.W. Main, Suite 500
Portland, Oregon 97204
(503) 275-9500

November 1989

TABLE OF CONTENTS

Introduction 1

Purpose For The Guide 1
Content Of The Guide 1
Types Of Instruments Included 1
Sources Searched 2

Issues In Assessing Speaking And Listening 3

What Skills Are To Be Measured 3
Issues Involving Tasks 4
Issues Involving Responses 8
Issues Involving Scoring and Reporting 9
Construct Validity 10
Ecological Validity 10
Reliability 11

Criteria For Describing, Selecting And Reviewing
Assessment Tools In Speaking And Listening -- Summary 13

Criterion 1: Content 13
Criterion 2: Reliability 14
Criterion 3: Validity 14
Criterion 4: Help With Interpretation And Use 15

Current State-Of-The-Art And Future Trends 17

Current State-Of-The-Art 17
Advice To Consumers 18
Future Trends 18

APPENDIX A: Sources Searched 21
APPENDIX B: Long Reviews 27
APPENDIX C: Short Reviews 59
APPENDIX D: Resources 93
APPENDIX E: Summary Table 97

Bibliography 119

Glossary 125

Index 131

INTRODUCTION

Purpose For The Guide

Speaking and listening are important both in school and in everyday life. Studies during the last 70 years have shown that students spend anywhere from 45% to 70% of their school day speaking and listening and that in daily activities people spend anywhere from 30% to 65% of their time in communication activities of which a large portion is listening. Listening is the first language skill we develop (followed by speaking), and our ability to read, write and learn from discussion contexts is directly related to our ability to listen and speak. Adequate oral communication frequently determines an individual's educational, social and vocational success. (Carbol, 1986; Ohio Department of Education, 1985; Plattor, 1988; Rubin and Mead, 1984; Wolvin, 1985)

Currently, many educational agencies are attempting to improve their attention to speaking and listening in curriculum and instruction. The purpose of this Guide is to assist educators to become more knowledgeable about tools for assessing speaking and listening. The Guide is designed for those somewhat knowledgeable in the areas of assessment and language arts instruction.

Content Of The Guide

This Consumer Guide has three major sections. The first is a short discussion of assessing speaking and listening. It includes information about definitions, taxonomies of skills, issues in validity and reliability, the current state-of-the-art in assessing speaking and listening, and what to consider when selecting an assessment tool.

The second section contains descriptions and reviews of assessment tools. Longer reviews are provided for instruments that are readily accessible, that measure some aspect of "communication competence" (defined below), and that have some technical information available. These longer reviews include descriptions of the purpose(s) the author sees for the assessment instrument; content (tasks, response modes and scoring procedures); reliability; validity studies; amount of help with interpretation and use; and source. These reviews are evaluative and attempt to relate each instrument to the issues and taxonomy framework discussed in the first section of the Guide.

Short reviews are presented for research instruments, achievement test series, instruments developed by educational agencies and instruments lacking technical investigation.

The final section lists additional resources available to the user such as print resources and professional organizations.

The Guide also includes a summary table of all instruments reviewed and an index so that instruments can be easily located.

Types Of Instruments Included

The primary focus of this Guide is published assessment tools designed for use in formal assessment settings. However, we have also included information about more informal instruments that could be used at the classroom level.

There are examples of assessment tools in various formats -- multiple-choice, observation, self-evaluation, peer-evaluation, and performance.

The emphasis is on instruments designed to measure some aspect of communication competence, defined as the ability to use communication to achieve a goal within a social context (Larson, 1978; Reed, 1984; Rubin, 1982; Wilkinson, et al., 1979). Less emphasis is given to instruments designed primarily to look at other aspects of speaking and listening such as physiological integrity (auditory acuity, speech defects, etc.), and linguistic competence, defined as the tacit knowledge required to form correct language, e.g., syntax, grammar and vocabulary (Larson, 1978). Accordingly, because of their emphasis on physiology and linguistic competence, we have excluded many instruments designed primarily for special education populations and ESL students.

Sources Searched

A complete list of sources searched to find instruments is provided in Appendix A. Briefly, these included ERIC, Rubin and Mead (1984), and Fagan, et al. (1985); publications of various professional organizations; all government-funded Labs and Centers; all state departments of education in the U.S. and Canada; listings from Buros Institute (Conoley and Kramer, 1989; Conoley et al., 1988; Mitchell, 1985; Buros, 1978), ETS' test collection, and test publishers' catalogs; professional journals; and experts at a number of colleges and universities.

ISSUES IN ASSESSING SPEAKING AND LISTENING

Both developers and users of assessment tools have a role in ensuring good and fair assessment (Joint Committee On Testing Practices, 1988). Developers need to conduct the studies and provide the information needed to enable users to select appropriate tests and interpret scores correctly. Users need to know their own purposes for assessment, select instruments that satisfy these purposes and are appropriate for the intended population, and interpret and use results properly.

The reliability, validity and usability issues that need to be considered when developing, selecting and using assessment instruments in speaking and listening will be addressed in the following broad areas:

A. Issues relating to the skills to be measured

B. Issues relating to the task presented to the student during the assessment

C. Issues relating to the responses that students make to the task

D. Issues relating to how responses are rated/scored

E. Issues in construct validity

F. Issues in how assessment tools relate to use

G. Reliability

The discussion is intended to describe the type of information in each area that developers should provide to enable users to judge whether the instrument is appropriate to their purpose and situation. These considerations are also used in the appendices of this Guide to describe and discuss the instruments reviewed. Thus, as we discuss the issues we will also indicate how this information will be used to describe, categorize and evaluate instruments in this Guide.

A. What Skills Are To Be Measured?

1. **Level Of Skills.** The skills necessary for effective listening and speaking are described differently by different people. Effective listening and speaking skills could include everything from the ability to articulate and hear properly to the ability to accomplish a purpose within a social context (Lundsteen, 1979; Barker, 1984; Wolvin, 1985). For ease in describing the different types of instruments available, we place them in three categories or levels:

   a. **Physiological.** These instruments measure the person's ability to hear and speak, e.g., auditory acuity and articulation. These instruments are outside the scope of this review.

   b. **Linguistic Competence.** These assessment tools tend to look at the sophistication of students with respect to the complexity of language they can produce and understand. These instruments cover such things as receptive and expressive vocabulary, the complexity of grammatical constructions used by the student, the length of student sentences, the complexity of sentences that students can understand, etc. Since people tend to use language of increasing difficulty as they get older, these types of measures are often used to tell how sophisticated a student is for his or her age in order to place students in various programs or to plan instruction.

c. **Communication Competence.** At this level, we are interested in how well students can use aural and oral skills to accomplish a goal within a social context. This is what we usually think of when we consider someone's ability to communicate. Although a certain degree of linguistic competence is required to do this, other skills are also required; for example, altering the level of language used to fit the audience and setting.

Communication competence is what we are trying to create in students. Physiological and linguistic competence are enabling skills, but are not the goal, just as decoding skills in reading are important, but do not constitute ability to read. Therefore, measures that cover only linguistic competence are not measures of "general oral language ability."

This Guide focuses on tools for assessing communication competence. However, in some cases there is not a clear distinction between what skills would be assessed to demonstrate communication competence and those assessed to look at linguistic competence. For example, the ability to follow orally given directions, a communication competence, involves understanding messages of various levels of complexity, a linguistic competence. Also, at the lower grade levels, listening comprehension involves both linguistic and communication competence. Thus, although some instruments can be categorized as primarily emphasizing linguistic or primarily emphasizing communication competence, many have aspects of both, especially in the area of listening comprehension.

In the reviews, we will describe the degree to which each instrument emphasizes communication competence. With respect to listening comprehension, we will describe items that cover vocabulary and decoding the meaning of sentences as measuring linguistic competence, and items that require recall of important facts and making inferences about longer passages as measuring communication competence.

Although our primary focus is communication competence, a few short reviews of instruments primarily dealing with linguistic competence have been included for comparison purposes.

2. **Skills Taxonomies.** Even within communication competence not all instruments cover the same things. For example, some focus on individual skills such as identifying the main idea, organizing ideas, and distinguishing fact from opinion, while others attempt to measure more global abilities such as whether students know how and when to apply various skills.

   In the reviews we describe the skills that each instrument attempts to measure so the user can determine which might correspond most closely to what he or she might want to measure.

3. **Sampling.** No assessment instrument can cover all the skills and processes of interest. One needs to sample from the skill domains. The trick is to sample in such a way that the results are an adequate indication of student performance and the use of results does not contribute to restricting curriculum or instruction. In the reviews this will be discussed as part of validity.

B. Issues Involving Tasks

Introduction. Every assessment requires that the student do some task. Tasks have implicit or explicit settings, audiences, purposes, and content. Communication competence cannot be assessed outside the context in which it occurs, because what may be effective in one context may not be effective in another. For example, it is not always appropriate or most effective to use long, complex sentences, big words, or formal language. Likewise, what might be most effective for a discussion with the teacher on a grade you would like to have, might not be the most effective in a group discussion with peers, or in a casual conversation with friends.

In the reviews, we describe the setting, audience, purposes and content for communication stated or implied by each instrument. The categories used are a distillation of the points of view of several sources, especially Backlund (1982), Hutchinson, et al. (1987), Rubin and Mead (1984), Ohio Department of Education (1985), Iowa Department of Education (1986, 1989), University of the State of New York (1988), Backlund (1985), Barker (1984), and Wolvin (1985).

1. **Purposes.** Purposes for speaking and listening include such things as providing information to a person, expressing an opinion, describing an event, carrying out required social pleasantries, and recreation. Although these purposes have been categorized differently by different authors, we will use the scheme proposed by the University of the State of New York (1988). They propose that speaking and listening are used for the following purposes in school and everyday life:

   a. **Social Interaction.** This includes social conversations, social rituals, functional communication (e.g., taking messages, describing incidents), etc.

   b. **Transmitting Information and Understanding.** This covers acquiring, interpreting, applying and transmitting information; for example, following instructions, comprehending what is heard, speaking so that others understand and communicating nonverbally.

   c. **Analyzing and Evaluating Messages.** This includes listening critically to the messages of others and expressing one's own opinion.

   d. **Appreciation and Entertainment**. This involves listening and speaking for recreation and expressing oneself.

   The communication context implied in an assessment is not always entirely clear. For example, should the context be described from the developer's perspective or the perspective of the student taking the test? For example, in listening comprehension, should the context be that of the test taking situation or that of the individual listening passage? If the former, then the context is one-way communication with the audience being the teacher for the purpose of being evaluated. If the latter, then the context could be whatever the passage covers; for example, a simulated conversation in which the purpose is social interaction, the audience is the participants in the conversation and communication is two-way. In our reviews we try to take whichever perspective seems most reasonable.

   (Possible purposes for communication should not be confused with the purposes for assessment described above. Possible purposes for doing an assessment include selection of students, accountability, planning instruction, and recording student progress. This is how the results will be used. The purpose for communication implied by the test is an aspect of test content -- from the student's point of view, what purpose does the communication within the test serve? The student could be trying to convince someone of something, exchanging information, socializing, etc.)

2. **Setting.** The setting for a communication includes such things as (Rubin and Mead, 1984; Ohio, 1985):

   a. **Group size** -- one-to-one, small group, large group, mass media.

   b. **Formality of the occasion** -- more formal settings are presentations, lectures, and classrooms; less formal settings are discussions with friends and playground conversations.

   c. **Format** -- interactive communication in which speakers and listeners interact with each other (e.g., discussion, interview, debate, social conversation) *versus* one-way

communication in which the speakers and listeners have much less opportunity to interact (e.g., speeches, listening comprehension, drama).

d. **Preparation** -- impromptu or prepared.

3. **Audience.** The audience for a communication is the person or persons with whom one is interacting or toward whom the communication is directed. Audiences for students could include peers, parents, teachers, employers, younger children, siblings, etc.

4. **Content.** The content of a communication is that which participants communicate about. This could include cooking, politics, commercials, emergencies, interviews, directions, etc.

Assessment Issues. With respect to tasks, things that can get in the way of accurate measurement include situations that don't mirror real-life tasks, do not elicit true student reactions, or that introduce task requirements that are extraneous to the competencies being assessed, or don't reflect the range of skills involved in communication competence. Some important issues (Backlund, et al., 1980; Booth-Butterfield, 1986; Bostrum and Waldhard, 1988; Carbol, 1988;Faires, 1980; Mead, 1978; Phillips, 1980; Rubin and Mead, 1984; Stiggins, 1981) are listed below. In the reviews, instruments will be examined for their attention to these issues.

1. **Sampling From All Possible Contexts.** If the intent is to assess communication competence, then the tasks need to sample from the entire domain of speaking/listening purposes, settings, audiences and content that are relevant to students at various grade levels. For example, one cannot infer competence in the entire domain of ability to communicate from a listening comprehension test in which short passages are read to the student, students cannot take notes and cannot ask clarifying questions, and in which responses are only identified and not produced.

   Most current instruments do not attempt to sample from the entire domain, but only focus on a few component skills. Therefore, in the reviews, instruments are described with respect to whatever aspect of communication competence is covered.

2. **Artificial Tasks.** Assessment tasks are often artificial to one degree or another. For example, some speech exercises require students to present a three minute persuasive talk on an assigned topic. From the student's perspective, the audience may be the evaluator and the purpose may be to evaluate speech competence. Would students feel the same personal relevance they would if presenting such a speech out of personal commitment? Would the same skills be exhibited?

   Similarly, how well can listening to tapes reflect real-life activities? Listening to tapes for information restricts interaction, and eliminates the visual aspect of communication.

   Another example is that many testing situations require some degree of role-playing, as in a simulated interview. The lack of ability to role-play may be mistaken for the lack of ability to communicate effectively.

   A final example is the degree to which objective format tests (e.g., multiple-choice) simulate real contexts and require the same skills.

   Developers of instruments should demonstrate that the situation presented to the student is an adequate substitute for a real-life situation, and that the responses elicited are an adequate representation of the real behavior. As situations become more artificial, the need for such proof becomes greater. Even tasks that are performance-based (such as a simulated interview or a set-up discussion) should have such documentation.

3. **Skills In Isolation.** The task environment can determine whether one is assessing skills in isolation or observing how skills are used in concert to achieve a goal. An example of assessing skills in isolation is a listening test in which students listen to a short passage and then pick or state the main idea. An example of a listening exercise in which skills are used in concert to achieve a goal is when students have to take notes on a lecture. This requires students to not only identify the main idea, but also choose the most important information, make inferences and write things down so that they can facilitate later recall. It may also require students to monitor their own comprehension so that they know when to ask questions.

4. **Tasks That Require Speaking And Listening.** Speaking and listening assessments need to reflect the unique aspects of speaking and listening rather than just being made parallel to reading and writing assessments. Speaking and listening are different from reading and writing in that (Lundsteen, 1979; Backlund, et al., 1980; Rubin and Rafoth, 1986; McCroskey, 1986):

   a. They are real time -- listeners don't have much control over the rate of presentation of material; the speaker has to come up with the most appropriate language quickly; the ability to go back over information is more limited; and the speech record is more impermanent.

   b. They have an extra visual and aural component -- there are non-verbal cues and cues based on how something is said.

   c. They involve different social relationships -- speaking and listening are face to face activities, and thus are different in style (more concrete and personal language, more awareness of time, place and occasion); language (simpler vocabulary, greater density of ideas); and the need for social interaction.

   d. They are less linear than reading and writing in the sense that there are pauses, incomplete sentences, repetition, etc.

   Thus, tasks used in assessing speaking and listening need to be structured differently from those for reading and writing, and must emphasize sets of skills that are somewhat different. For example, Rubin and Rafoth (1986) propose that material to be presented and listened to orally must have certain characteristics if it is to be effective. They call these characteristics "listenability." Material is listenable when, for example, sentence structure is simple, passages contain a high degree of redundancy, thematic units are resolved quickly, and the language used is that of face-to-face interaction. Therefore, passages to be used in listening comprehension tests should not just be any written item presented orally, but must be listenable. Likewise, speakers could be rated on the extent to which their oral presentation is listenable.

   If assessment tasks are artificially set up so that these features are not present, the developer needs to provide proof that performance can be generalized to those situations in which these features are present.

5. **Individual Differences and Bias.** The task itself can produce inaccurate results for certain individuals or groups.

   a. Some topics might be more familiar to some students than others. This might, for example, enable one student to do better on an impromptu speech than another.

   b. Differences in communication anxiety between students might interact with the task to provide over- or underestimates of performance ability. For example, one study (Booth-Butterfield, 1986) showed that high anxiety students did better with more task structure, while low anxiety students did better with less task structure. McCroskey and Daly (1987)

includes several articles on how communication anxiety and other personality variables influence communication competence.

c. Presenting listening stimuli on tape (versus having the teacher read the material) may affect some students more than others (accent, anxiety, etc.)

d. Students may be differentially interested in passages to be listened to or topics to be spoken about. Also, some cultural groups might be more tolerant of materials they consider boring. Ting-Toomey and Korzenny (1989) present a wide-ranging treatment of language, communication and cultural relationships.

e. Some cultural groups may be less willing than others to speak orally, express opinions, and offer information unless they consider themselves expert. Similarly, children in various cultural groups respond differently to adult questioning.

f. Some people have better memories than others. Students who have compensated for this by learning to take notes, ask questions, etc., may be penalized by listening tests that require a high memory load.

g. How the task is presented to the student can affect performance. The student may misunderstand the task, the way the task is presented may not stimulate the retrieval of relevant skills, or the way the task is presented may stimulate anxiety on the part of the student.

Instruments should discuss these issues and provide information on the extent to which these things may be expected to occur.

C. Issues Involving Responses

Introduction. The types of responses required from students can have an effect on the assessment of their communication competence. Response requirements that are not realistic or that introduce the need for skills that are extraneous to the ones being assessed can get in the way of accurate assessment.

Students can demonstrate knowledge or skill by responding to written multiple-choice questions, pointing to pictures, making a presentation, having a discussion, evaluating themselves, evaluating peers, physically following an instruction, etc. In general, these activities can be placed in two categories -- objective format and performance. Objective format responses involve the *identification* of a correct answer. Performance responses include any format that requires the *production* of a response, for example, a short answer, a speech, or performance of some task.

Issues. Some issues with respect to responses are (Hohl and Cheney-Edwards, 1976; Rubin and Mead, 1984; Spandel, 1989; Stiggins, 1981):

1. **Objective Formats.** The advantages of objective formats are that they are very easy to give and score. (The argument that they are also easy to construct does not apply to assessing communication competence because adequate measurement of this area can be very tricky in an objective format.)

   Drawbacks are that they only have one right answer, they tend to assess skills in isolation, they are identification rather than production tasks, and they often do not present information in a manner that is seen by teachers as being useful in classroom situations.

It is possible to construct an objective format test that measures communication competence. However, the developer needs to provide evidence that performance on the test is an adequate representation of performance in real-life settings.

2. **Performance Formats.** The advantages of performance assessments are that the context and task often can be made more realistic; for example, actually giving a speech, participating in a discussion or taking notes while listening to a lecture. This can help to put responses in a context, promote skills working in concert, allow for more than one right answer, promote thinking skills, allow one to assess more types of skills, and assess how students actually use skills. Additionally, teachers often view the results as bearing more directly on what they do in the classroom.

   Disadvantages of performance assessments are that they are often more costly to give and score (in time, money and need for expertise and training); it is difficult and costly to sample an appropriate range of contexts and performance; and they are not immune from being artificial. If the latter is the case, then, as with objective-format tests, validity studies need to be done to show that results mirror performance in real-life.

3. **Extraneous Response Requirements.** Extraneous response requirements are those skills the students must use in order to respond in an assessment, but that have nothing to do with the skills being assessed. Some examples are the need to demonstrate speaking and listening competence through responses that require reading and writing; emphasis on standard English usage regardless of the purpose and context; the need to role-play; and test-wiseness. Inability in one of these areas might be mistaken for lack of communication competence.

D. Issues Involving Scoring And Rating

Issues. The issues discussed in this section relate mainly to performance assessments. Each performance must be judged by *someone* using *some* set of criteria. Issues include:

1. **Correspondence Between Criteria and Task.** The dimensions rated and the criteria for rating have to correspond to the task. For example, you would allocate more importance to "provides adequate support for an opinion" if the student were making a persuasive speech than if he or she were giving directions to someone younger.

   Some rating dimensions might hold across contexts. For example, in speaking, one might always include the general categories of language use, mechanics of delivery, content and organization. Even so, the specifics in each area to be considered when rating a performance will likely be different depending on the purpose, setting, audience and task.

2. **Subjective v. Objective Judgments.** Subjective approaches require someone's judgment as to the quality of a performance. Judgments can be holistic (overall impression), primary trait/focused holistic (whether the performance accomplishes its purpose), analytical (how the performance looks along various dimensions) or dichotomous (which specific things are present or absent).

   Objective approaches attempt to bypass subjectiveness in scoring. For example, in a persuasive speech one might look at how many listeners change their minds as the result of the speech. Or, to judge descriptive ability, one might have a speaker describe something to an audience and then see how well the audience can reproduce it. The problem with such approaches is that the outcome is as dependent on the abilities of the audience as it is on the abilities of the speaker. Currently, it seems that more direct assessments by trained raters are better for getting at the desired performance.

3. **Rater Effects.** Raters can produce inconsistent ratings for a number of reasons. They may have a different understanding of the criteria to be used, dislike of specific things such as behaviors or word choices, bias toward various groups, etc. Raters need to be carefully trained. Instruments requiring ratings of student performance should include procedures for training, detailed descriptions of scoring rubrics, and sample student "anchor performances" that illustrate the various ratings.

4. **Rating From Memory.** Some procedures may require that teachers rate students based on their memory of general student performance in the classroom. These ratings can be very unreliable (Massachusetts, 1982; Arter, et al. 1986). Memory ratings can be useful for informal, classroom assessment, but when used for formal purposes, they should be done with proper training, and even then with great care.

E. Construct Validity

Many of the issues discussed above relate to the need to demonstrate construct validity -- that an instrument measures what is claimed. This can be difficult because of the lack of independent criteria for establishing communication competence. In establishing, for example, how well an artificial performance task elicits a "real" behavior, the only way to discover what that "real" behavior is is to make further observations and judgments. No outside, objective procedure exists. Therefore, the only way to establish validity is through a series of studies in which the instrument provides results that support the inferences to be made from the test scores.

Such studies typically address such things as whether performance on a task improves with age and/or training; whether performance reflects expected differences between existing known groups; the degree of correlation between the instrument being tested and other instruments that purport to measure the same thing; the relationship between test results and other assessments of classroom work (e.g., teacher judgments, grades, detailed observations); subjective judgments by experts and teachers that the instrument measures what is claimed; how differences in student background knowledge, communication anxiety, interest, gender, ethnic group and memory affects performance; how differences in task presentation or response requirements affects performance; student perceptions of the realism of the task; and whether content is based on a model or theory of communication.

In our reviews, instruments will be rated on the quality of validity studies.

F. Ecological Validity

An instrument is ecologically valid when (Tittle, 1989):

1. The assessment tool is used and interpreted properly. This means that users understand the scores and do not use the assessment tool for purposes not supported by available validity information. If relevant evidence is not supplied by the developers, users should obtain it themselves.

2. The results are perceived as being useful and are actually used. Assessment results can be used and interpreted properly and still not be perceived as being useful. Likewise, assessment results can be perceived as being useful and not be used.

3. Use of the tool does not promote negative effects such as restricting the curriculum or encouraging students to focus on certain skills to the exclusion of others. Assessment instruments assess a subset of skills from a broad domain. Problems can occur when so much importance is placed on test results that only the subset of skills assessed by the instrument is included in instruction.

Such restriction can occur with both objective-format tests and performance assessments. In the area of performance assessment, for example, if the assessment only requires students to speak persuasively, teachers might focus on these sets of tasks to the exclusion of group discussion, interactive communication, personal expression and speaking for other purposes.

4. There are direct links to instruction.

Assessment materials should provide information that allows users to select instruments that meet their needs and that help with proper interpretation and use of results (Joint Committee On Testing Practices, 1988). It is also desirable that the other aspects of ecological validity discussed above be addressed. In the reviews, each instrument is rated on these criteria.

G. Reliability

The reliability of an assessment tool includes the degree to which results are accurate or replicable across forms (alternative form reliability), occasions (test-retest reliability), and raters (interrater reliability). In addition, internal consistency reliability refers to how well the test samples from a single dimension.

In instruments that use ratings, interrater and test-retest reliability are the most important types. Extensive training of raters is often required to obtain ratings that are consistent both across raters and within the same rater over time.

For objective format tests, test-retest and internal consistency reliability are the most frequently used. Alternate-form reliability applies to both performance and objective format tests when there is more than one form.

The degree of reliability required varies with the intended use of the test. Tests to be used for very important and difficult to reverse decisions about students (such as promotion and graduation) need to have reliabilities above .95. Tests to be used informally for easy to reverse decisions don't need to have such high reliability (although good reliability is always an asset.) For such tests, reliabilities in the range of .70 and above may be adequate.

When evaluating reliability it is also important to consider the group on which the reliability coefficient was calculated. Reliabilities need to be calculated on the same types of students under the same circumstances as those of the user.

The specific criteria used in this Guide to rate the reliability of instruments is presented in the next chapter.

CRITERIA FOR DESCRIBING, SELECTING AND REVIEWING ASSESSMENT TOOLS IN SPEAKING AND LISTENING -- SUMMARY

This section summarizes the considerations presented in the previous chapter and provides the criteria used in this Guide for rating assessment tools.

To use this guide effectively, it is paramount that the user know the purposes for which his or her assessment results will be used, for these affect the qualities one must look for in an instrument. The two examples below illustrate this point.

1. In general, the more important (and less reversible) the decision made about the student, the greater the requirements in terms of formal development, training, and proof of technical quality. When the purpose is classroom assessment, for example, there is not as great a requirement for proof of technical rigor as is the case with a large-scale, high stakes assessment. This is because for classroom assessment one might be more interested in a broad array of approaches and their relationship to instruction than in proof of technical rigor; because there are many pieces of information available to modify the conclusion drawn from any single piece; and because, for instruction, it is sometimes better to have information that is broader but less accurate than information that is highly accurate but restricted in scope.

2. The desired content of the assessment can vary according to purpose. For example, a minimum competency test would focus more narrowly on skills that are defined as essential and would measure at the level of difficulty that is considered minimum for effective functioning. An achievement test would require broad coverage and enough ceiling and floor to effectively measure students at various levels of achievement. A diagnostic test might have thorough coverage of a narrower range of skills and prerequisite subskills.

To be of service to users with a broad range of needs, we will be as descriptive as possible about the content, tasks, contexts and technical attributes of instruments so that the user can decide the extent to which an instrument will match their purpose. The criteria presented below for rating instruments are meant to be suggestive of typical criteria for looking at instruments. The user might alter these criteria, or the weight given any one, depending on his or her purpose for assessment.

Criterion 1: Content

We will describe:

1. The **purposes/uses** the author planned for the instrument.

2. **General information** about the instrument such as the grade levels intended for use, number of levels, forms and items, test length, and administration requirements (training, equipment, etc.).

3. The **task** presented to the student, including the purpose, setting and audience for the communication, as well as the specific content presented to students and the skills the assessment is trying to cover. With respect to skills, we will indicate both the extent to which the assessment tool emphasizes linguistic versus communication competence and the specific skills covered.

4. The **responses** by which the student demonstrates his or her level of skill.

5. Who **scores** the responses or performances and the criteria by which they are scored.

The rating in this area will depend on how well materials accompanying the instrument provide the information necessary for users to match the instrument to their needs.

Excellent	The developer includes information on purposes, the population recommended for use, and limitations of the instrument for the use suggested; describes how the instrument could be used with atypical populations; defines measurement terms and uses language appropriate for the user; lists specialized skills needed to administer the instrument; describes the test development process; provides information on reliability and validity; and provides samples of questions, directions, answer sheets, manuals and score reports (Joint Committee On Testing Practices, 1988).
Good	Much of the information above is provided.
Fair	Some of the information above is provided.
Poor	Little of the information above is provided.

Criterion 2: Reliability

We will use the following criteria for judging the general adequacy of the reliability of instruments:

Excellent	Reliability of total test score .95 or above; reliabilities of subtest scores .90 or above.
Good	Reliability of total test score .85-.94; reliabilities of subtest scores .80 and above.
Fair	Reliability of total test score .75-.84; reliabilities of subtest scores .65 and above.
Poor	Reliability of total test score .74 or below; reliabilities of some subtest scores below .64.
Unknown	No information is provided.

Criterion 3: Validity

In the reviews of instruments, we describe the types of validity considerations and studies carried out by the author(s). This includes discussions of content, criterion and construct validity. Because they relate most directly to speaking and listening, we will pay particular attention to the validity issues discussed in the previous chapter: extent of sampling from contexts, artificial v. naturalistic tasks, assessing skills in isolation or in concert, tasks that require extraneous skills, sources of bias, degree of realism in the task and response, extraneous skills required for responding, correspondence between the task and scoring criteria, rater effects, and ecological validity.

For purposes of this Guide, ratings in the area of validity will be:

Excellent	There are many lines of evidence presented that the instrument measures what is claimed and can be used for the purposes proposed.
Good	Several lines of evidence are presented and these provide convincing evidence.
Fair	At least one study was completed and this provides convincing evidence.
Poor	Evidence that is provided is not convincing.
Unknown	No evidence is provided.

Criterion 4: Help With Interpretation and Use

Ratings in the area are:

Excellent	There are norms that are based on a large, representative sample of an appropriate reference group of students or there are other useful standards for comparison (e.g., performance of various groups or judgments of mastery); there is help in how to use the results in instruction; there is a discussion of the possible uses and misuses of results; there are good score reports and they serve the intended use.
Good	There are appropriate norms and/or other standards of comparison. There is discussion in at least one other area mentioned above.
Fair	There is good assistance in at least one of the areas mentioned above.
Poor	The assistance that is provided is judged seriously lacking.
Unknown	No information is provided.

CURRENT STATE-OF-THE-ART AND FUTURE TRENDS

Current State-Of-The-Art

Currently, many of the assessment devices labelled "oral language," "language," "listening comprehension," "language ability," "oral communication," etc. measure linguistic competence. These tend to cover such things as receptive and expressive vocabulary, understanding or producing sentences of increasing complexity, understanding the referents of pronouns, etc. They also tend to measure isolated skills that are largely without specified or implied contexts.

In addition, many of the instruments that claim to measure "language ability" do not define what is meant, and it is therefore easy to infer more from the results of these instruments than is warranted. Thus, from the titles, one cannot necessarily differentiate those which focus on linguistic competence and those which cover some aspect of communication competence. Even the instruments calling themselves "listening comprehension" cover anything from auditory discrimination to finding the main idea. There seems to be no common consensus on what should be included on an oral/aural language instrument even though most developers claim to have consulted experts, reviewed the research literature and/or reviewed the most common curriculum materials.

Of the assessment devices that measure some aspect of communication competence in listening, most emphasize listening comprehension -- a mix of linguistic competence and communication competence. These typically entail listening to passages of varying lengths and discourse modes, and answering a variety of multiple-choice questions about the passages. In most achievement test series, both the passages and questions are read to the student by the teacher. In many state assessments and other individually prepared assessments, the passages and questions are provided on tape. The tests usually measure isolated skills, although there is sometimes an attempt to put them in context. The testing situation usually entails short passages, use of formal English, one-way communication (students cannot ask questions), and varying amounts of memory load (students almost always cannot take notes or listen to the passages again).

There are only a few listening instruments that attempt to look at interactive speaking/listening, other purposes for listening besides transfer of information, interpreting nonverbal cues, using inflection and intonation to interpret meaning, or listening in naturalistic settings. There are only a few that incorporate assessment approaches other than multiple-choice. We found almost no assessment devices for looking at how well students use various strategies for listening effectively. Most studies of validity do not entail attempts to see how performance on the test relates to ability to communicate in daily life.

In speaking, most of the instruments that attempt to look at some aspect of communication competence focus on extended monologues (simulated speeches) in various modes (narrative, persuasive, expository, etc.) with analytical ratings done by the classroom teacher. There are also a number of informal teacher checklists, peer-ratings, and self-ratings for speaking. There are a few rating forms for looking at group discussions and social interactions. Most speaking assessment tools are appropriate mainly for informal classroom use, although there are a few examples of standardized performance assessments similar to those developed for writing. Most speaking instruments entail artificial (versus naturalistic) tasks.

There is still much lacking in the area of assessing communication competence. The skills and contexts covered by most current assessment devices are restricted to those easiest to measure. In addition, validity studies typically do not attempt to see how performance on the instrument relates to everyday ability to communicate.

The best overall instruments we reviewed were the English Language Skills Profile (speaking and listening), the PONS (nonverbal communication), the Communication Competency Assessment Instrument (speaking and listening) and the College Outcomes Measures Project (speaking and listening).

In the area of standardized, multiple-choice instruments of listening comprehension, the better tests were the CAT/CTBS listening supplement (see Listening Test), the ITBS/TAP listening supplements (see ITBS and TAP Listening Supplement), the SAT, the National Achievement Test and the Survey of Basic Skills. Please remember, however, that many of these could benefit from additional theoretical justification and study of validity.

Advice To Consumers

Based on our review of current instruments, our advice to consumers includes:

1. Consumers should be clear on what they want to measure so that they can find an adequate match with an assessment tool. Clarity includes definitions, theoretical position, and how various student skills could manifest themselves. Consumers need some expertise in the area of speaking and listening so that they can adequately decide how well an assessment tool covers material desirable for their own purposes. There is some good material available, but it needs to match with user needs.

2. Consumers should be clear on their purposes for assessment.

3. Don't trust titles. Look at the actual content of the test in addition to the author's descriptions of what the test measures and why.

4. If results are to be used for formal purposes, consumers should be prepared to assess the validity of instruments for these uses.

Future Trends

There are several trends in the current literature on how to best assess speaking and listening. These include:

1. More emphasis on communication competence in addition to linguistic competence

2. Attempts to identify a broader array of communication contexts and sample from this array so that results are more comprehensive and representative

3. Attempts to assess interactive communication rather than just one-way communication

4. The advocacy of integrating speaking and listening skills with other communication skills and with content areas

5. More consciousness of the effects of task and cultural context on performance and the need to match rating with context

6. The advocacy of the use of portfolios in order to gather a variety of information from a variety of sources

7. More awareness of the things that can make an assessment invalid -- task conditions, response conditions, scoring, previous knowledge, etc.

8. More attention to the effect of personality on communication competence

Assessment devices designed around these considerations will provide an advance in the field. They will, however, have to be accompanied by appropriate validity studies to show that they do provide an adequate estimate of general "oral language proficiency" in real life.

APPENDIX A

Sources Searched

Bibliographies

Brody, C. (1986) A guide to published tests of writing proficiency (Second edition). Northwest Regional Educational Laboratory, 101 S.W. Main St., Portland, Oregon, 97204.

ETS Test Collection (1984). Composition and writing skills. Educational Testing Service, Princeton, NJ.

Fagan, W.T., Cooper, C.R., and Jensen, J.M. (1985) Measures for research and evaluation in the English language arts, Volume 2. National Council of Teachers of English, 1111 Kenyon Road, urbana, Illinois 61801.

Goodman, K.S., Goodman, Y.M., and Hood, W.J. (1989) The whole language evaluation book. Portsmouth, NH: Heinemann Press.

Hammill, D.D., Brown, L., and Bryant, B.R. (1989) A consumer's guide to tests in print. Austin, Texas: Pro-Ed.

Hepner, J.C. (1988) ETC Test collection cumulative index to tests in microfiche, 1975-1987. Princeton, NJ: Educational Testing Service.

Mitchell, J.V. (1985) The ninth mental measurements yearbook (Buros). Lincoln, Nebraska: The University of Nebraska Press.

Rubin, D.L. and Mead, N. A. (1984) Large scale assessment of oral communication skills: Kindergarten through grade 12. ERIC 245 293. Also Speech Communication Association, 5105 Backlick Road, Annandale, Virginia 22003.

Test Collection, Educational Testing Service. (1986) The ETS test collection catalog, Volume 1: Achievement tests and measurement devices, Phoenix, Arizona: Oryx Press.

Publisher's Catalogs

Academic Therapy
American College Testing Program
American Testronics
CTB-Mcgraw Hill
Curriculum Associates
Emporia State University
Educational Records Bureau
Educational Testing Service
GED
Instructional Objectives Exchange
Institute for Personality; and Ability Testing (IPAT)
National Assessment of Educational Progress (NAEP)
New Zealand Center For Education Research
NFER-Nelson (England)
Psychological Assessment Resources (PAR)
Pro-Ed
Psychological Corporation
Publisher's Test Service
Riverside

Science Research Associates (SRA)
Scholastic Testing Service
Slosson
Spectra
United Educational Services

Educational Agencies

Letters were sent to all state departments of education, U.S. territories and Canadian provinces. We have materials from 36 states, five provinces and one territory. These are: Arizona, California, Connecticut, Florida, Georgia, Hawaii, Illinois, Indiana, Iowa, Kansas, Kentucky, Maine, Maryland, Massachussetts, Michigan, Minnesota, Missouri, Montana, Nebraska, New Jersey, New York, North Carolina, Ohio, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming, Puerto Rico, Alberta, British Columbia, Manitoba, Northwest Territories, Ontario and Saskatchewan.

We also sent special request letters to some educational agencies when we knew about something in particular they were working on. Special letters were sent to California, New Hampshire, Illinois, Vermont, Connecticut, Michigan, British Columbia, Rochester Public Schools (NY), Salem-Keiser School District (OR), and Valley Education Consortium (OR).

Professional Organizations

American Speech, Language, Hearing Association
International Listening Association
International Reading Association
Conference on College Composition and Curriculum
National Council of Teachers of English
Speech Communication Association

Specific Individuals

Behnke, R., Texas Christian
Breadloaf School of English at Middlebury College, VT
Cooper, C., U. CA, San Diego
Graves, D. H., Writing Process Lab, U. of NH
Haas, C., Carnegie-Mellon
Hamp-Lyons, E., English Composition Board, U of MI, Ann Arbor
Lucas, C., San Francisco State
Murphy, S., San Francisco State
Quellmalz, E., RMC Reserach Corp., Mountain View, CA
Smith, M.A., Bay Area Writing Project

Labs and Centers

Appalachia Lab, Charleston, WV
Far West Lab, San Francisco, CA
Mid-Continent Lab, Kansas City, MO
North Central Lab, Elmhurst, IL
Northeastern Lab, Andover, MA
Research For Better Schools, Philadelphia, PA
SEDL, Austin, TX
Southeast Lab, Research Triangle Park, NC
SWREL, Los Alamitos, CA

Center for Applied Linguistics
Center for Bilingual Research, Los Angeles, CA
Center for Effective Secondary Schools, Madison, WI
Center for Research In Vocational Education
Center for Research to Improve Postsecondary Teaching, U of MI, Ann Arbor
Center for Social Organization of Schools, Baltimore, MD
Center for the Study of Evaluation, UCLA
Center for the Study of Reading, U of IL
Center for the Study of Writing, UC Berkeley
Learning, Research and Development Center, Pittsburgh, PA

APPENDIX B

Long Reviews

INTRODUCTION

Long reviews are provided for instruments that:

1. Are readily available
2. Are intended for commercial use
3. Have some feature of interest, such as extra technical information, speaking subtests or good content coverage

The best instruments in terms of available technical information and content coverage are included in this section, although not all the instruments in this section are rated as being the best.

Title:

CIRCUS, 1976.

Author(s):

Scarvia Anderson and Gerry Ann Bogatz

Source:

CTB/McGraw-Hill, 2500 Garden Road, Monterey, California 93940, (800) 538-9547.

Authors' Description of Purposes:

"The CIRCUS program is based on the premise that a child's development has many dimensions and to truly understand his other educational needs, a variety of different abilities and skills needs to be evaluated. CIRCUS may be used in several ways, including program evaluation... individual assessment... and pretesting and postesting." (Manual, p. 4)

Authors' Description of Subtests:

(Note: Only the subtests relating to speaking and listening skills are included.)

Listen To The Story:	Level A measure simple comprehension of what is said and more complex interpretations. Level B assesses children's ability to comprehend and interpret oral language, but also incorporates receptive vocabulary and aspects of functional language.
Listening:	The listening tests (Levels C and D) measure the child's ability to listen to a narrative, understand and interpret events in it, remember the sequence of events, and understand vocabulary.
Say and Tell:	Say and Tell attempts to provide a reasonable sample of the richness of the child's oral language. Say and Tell has three parts -- A description of an object, ability to use different forms of words and a narrative.

Description:

The CIRCUS has four levels covering grades PreK-3. There is one form at each level. (A few of the subtests have two forms per level.) Subtests include both listening and speaking.

The listening subtests (called "Listen to the Story" at Levels A and B, and "Listening" at Levels C and D) require matching a picture to a sentence and marking a picture that answers questions about a narrative passage -- sequence of events, inferences, recall of information, and vocabulary. These questions tap both linguistic and communication competence. The passage at each level is an ongoing narrative about a circus. The narrative is stopped each sentence or two to ask a question. The teacher reads all passages and questions. This is a group test. Numbers of items are: Level A (25), B (36), C (40) and D (40). The publisher estimates that the test takes 30-40 minutes to give, depending on level.

Say and Tell (Levels B-D) has three parts, administered individually. Part 1 requires children to describe objects. The first object is described through oral responses to questions posed by the teacher, e.g., "What color is this?" This is scored on a three point scale depending on accuracy of the response. The second object requires a free response to the question "Tell me about what you

have in your hand." The teacher rates the description in terms of whether or not the following were included -- label, color, shape, material, primary function, design and sensory aspects.

Part 2 in Say and Tell assesses the child's ability to use plurals, verb tenses, prepositions, subject-verb agreement, comparatives, and possessives. The exact coverage depends on level. An example of the type of items is: A statement is made about one of two drawings and the child is asked to complete a statement about the other, e.g. "Here is a tree. Here are two _____." Sometimes students also provide a short answer to what is happening in a picture or complete a sentence about a picture. Children receive a score of 1 to 3 depending on the correctness of their response.

In Part 3, children describe a picture. The child's story is written verbatim and then scored for total number of words used, number of different words used, the complexity of sentences used and some aspects of the quality of the response.

The publishers estimate that Say and Tell takes about 20 minutes per student to administer and score. Say and Tell seems to measure mostly linguistic competence (e.g., vocabulary, complexity of sentences and knowledge of grammar), with some aspects of communication competence (e.g., quality of responses and inferences). Scoring is assisted by detailed charts of various responses and the ratings assigned to each.

Level A has three extra listening subtests: What Words Mean (receptive vocabulary -- 40 items, taking about 30 minutes), How Words Work (understanding sentences that emphasize syntax, word order and vocabulary -- 26 items, taking about 25 minutes), and Noises (identifying real-life sounds presented on tape -- 24 items, taking about 30 minutes).

Purposes:	Transmitting information
Setting:	Classroom, one-to-one, formal, one-way communication
Audience:	Teacher
Responses:	Multiple-choice, short answer, performance, impromptu, skills in isolation and skills in concert
Level:	Linguistic and communication competence

We rate the manual as "fair" - "good" in terms of the information necessary to select a test. The main problem is the lack of description of the theoretical basis for what the listening and speaking subtests are trying to accomplish, and therefore what inferences can really be made about the results.

Reliability:

Internal consistency reliabilities for the speaking and listening subtests range from .49 to .90. Most are in the upper .70's and .80's. These are rated as "fair" to "good."

Validity:

The CIRCUS was originally developed by ETS. Development was based on:

1. Sampling from those aspects of student performance in the early school years that were important for teachers to understand about children, and that could be most readily affected by instruction. This was determined by a survey of educational practices and curriculum materials.

2. Pilot-testing items. Item statistics are available.

3. Bias studies for ethnicity and gender.

4. A factor analysis to see how the various subtests relate to each other. The listening subtests tended to show a considerable effect due to a single underlying factor, proposed to be general ability. This raises the question of how well the listening test measures features unique to listening. The speaking subtests correlated less highly with the other subtests, indicating that it measures something somewhat different, as expected.

5. Relationships between listening and teachers' ratings is moderate. Such relationships with the speaking subtest are much lower. The publishers propose that this might be due to a ceiling on the test or teacher inconsistencies in rating oral language production.

This evidence of validity is rated as "fair".

Help With Interpretation:

There are several types of information provided to assist with interpreting scores. There are norms based on a large national population. Other standards of comparison are average scores of various groups of students in the norming sample, the percent of children in the norming sample getting each item right, and group norms tables. This information was originally developed between 1972 and 1977 and has not been updated. There are no plans at this point to update the test or norms. Therefore, these standards of comparison are somewhat outdated.

There is also assistance with interpreting results including how to develop local norms, profiling student and group proficiency, tracing progress over time, what the different scores mean, expected growth for students at various score levels, and appropriate cautions. However, this assistance relates to all the subtests in general. Specific assistance with interpretation and use of the oral language subtests is lacking.

The rating is "fair" mainly because of the age of the norms and lack of special help with oral language. The other assistance is good.

Comments:

The CIRCUS covers some aspects of listening and speaking in the lower grades. It includes performance measures. It's major drawback is the age of the norms (and other statistical information). There should also be additional validity work to show how well performance on the test relates to real-life communication success.

The CIRCUS received two moderately positive reviews in Buros Mental Measurement Yearbook (Mitchell, 1985, 9:224). Rubin and Mead (1984) concluded that the listening test "is a well designed test with a rigorous research base for assessing general school readiness. It is *not* a test of speech communication ability but a paper and pencil test. Listening measured in this way correlates with reading ability. The relationship to the ability to talk with or inform others is unknown" (p.32). They judged the speaking subtest to be "an adequate sample of children's productive language" (p. 34).

Title:

The Communication Competency Assessment Instrument (CCAI), 1982

Author:

Rebecca Rubin

Source:

Speech Communication Association, 5105 Backlick Rd., Annandale, VA 22003.

(Note: The description below is based on several studies using the test. We were not aware of the published source until just before publication. However, the information below has been reviewed by the author.)

Author's Description of Purpose:

"The CCAI was developed as a comprehensive college-level communication competency measure. The goal of the instrument was to identify students who may have difficulties with both sending and receiving communication in an educational setting" (Rubin and Roberts, 1987).

Author's Description Of Subtests:

The CCAI provides ratings on 19 communication competencies: pronunciation, facial expression/tone of voice, articulation, persuasiveness, and clarity of ideas; ability to express and defend a viewpoint, recognize misunderstanding, distinguish fact from opinion, understand suggestions for improvement, identify instructions, summarize, introduce self to others, obtain information, answer questions, express feelings, organize messages, give accurate directions, describe another person's viewpoint, and describe differences in opinion.

Description:

The CCAI was developed for use with college students but could also be used in high school. There is one form and one level. The assessment has three parts. The first task asks the student to present a three-minute extemporaneous persuasive talk on a topic of interest to the student. The performance is scored analytically on pronunciation, facial expression/tone of voice, speech clarity, informative/persuasive distinction, clarity of ideas, and ability to express and defend a point of view. An additional question assesses the student's ability to recognize a lack of understanding in the audience.

The second task requires students to watch a videotaped seven minute, forty second class lecture in which the instructor explains course requirements, explains factors that affect listening, gives suggestions for improvement and gives the first class assignment. The student then responds verbally to four questions about the lecture. These assess the ability to differentiate between fact and opinion, understand suggestions, identify the work needed to complete an assignment and summarize.

The final task requires students to respond verbally to statements about experiences he/she has had in an educational environment. Responses are evaluated in terms of ability to introduce oneself, ask questions, answer questions, express feelings, use a topical order, give accurate directions, describe another's viewpoint and describe differences in opinion.

The test is individually administered and all responses are verbal and open-ended. The CCAI takes about 30 minutes per student to administer. Ratings that would be considered "passing" are included.

Using our descriptive scheme, the CCAI can be described as:

Purposes:	Social interaction, transmitting information, analyzing messages
Setting:	One-to-one, formal and informal language, interactive and one-way communication
Content:	Artificial, persuasive, expository
Audience:	Assessor
Responses:	Performance, skills in concert, impromptu
Level:	Communication competence

It is not possible for us to rate the manual in terms of the information provided, because at the time of publication we had not obtained the manual.

Reliability:

Information about reliability is available from a number of sources. In these studies, interrater reliabilities range from .83 - .97. Internal consistency reliabilities range from .78 - .86. These are rated as "fair" to "excellent."

Validity:

A number of studies have used this instrument. Information includes:

1. The instrument was based on a review of current instruments and guidelines published by the Speech Communication Association.

2. The instrument was pilot-tested several times and reviewed by the communication faculty at the university.

3. The listening portion was moderately related to other tests of listening comprehension.

4. Correlations with other measures of student functioning -- ACT Ennglish scores, high school speech communication courses, persuasive speaking grades, credits completed, GPA, communication courses completed, teacher ratings and speaking experience) are low to moderate. The patterns of correlations are also as expected; for example, scores are related to judgment of competence but not to composure.

5. Certain combinations of item scores are highly effective in correctly placing student teachers in competency groupings.

6. There are moderate negative correlations with communication apprehension which shows that performance can be affected by student anxiety.

This evidence is rated as "good."

Help With Interpretation And Use:

This cannot be rated because the manual was not obtained as of the time of publication.

Comments:

This instrument is of interest because of it's performance orientation and it's attempt to sample from communication contexts needed for effective college classroom functioning. Training would be needed to adequately rate students.

A review by Spitzberg (1988) concludes that "despite a substantial amount of work done on the CCAI there are still questions that need to be addressed....It still remains to be seen whether or not these competencies make a 'real difference' outside the academic setting....In addition, several of the stimulus prompts may not be assessing *ability to perform* so much as the subject's comprehension of the prompts."

Other references include: Rubin (1982, 1985), Rubin and Graham (1986), Rubin and Feezel (1986), Rubin and Roberts (1987).

Title:

College Outcome Measures Program (COMP), 1983 - 1986

Authors/Source:

College Outcome Measures Program, The American College Testing Program, P.O. Box 168, Iowa City, Iowa 52243.

Authors' Description of Purpose:

"The College Outcome Measures Program (COMP) can help you focus on the development of the knowledge and skills acquired in general education courses...to meet a variety of goals. For example, you can use COMP to help reshape your curricula or design more effective learning activities...With COMP you can also help students use existing general education courses and programs in ways that will best enable them to achieve their personal and professional goals. COMP can help you determine whether students are reaching general education goals and whether they are receiving recognition for doing so. COMP can also assist you in communicating the value of general education to students, parents and other publics." (COMP brochure, p. 3)

"The COMP...(helps) you assess the extent to which your students are acquiring the knowledge and skills that characterize broad-based learning." (COMP brochure, p. 4)

Authors' Description of Content:

The areas measured by the COMP are:

Communicating:	Can send and receive information in a variety of modes (written, graphic, oral, numeric and symbolic/nonverbal), within a variety of settings (one-to-one and in small and large groups), and for a variety of purposes (for example, to inform, to understand, to persuade and to analyze).
Solving Problems:	Can analyze a variety of problems (for example, scientific, social and personal); select or create solutions to problems; and implement solutions.
Clarifying Values:	Can identify one's personal values and the personal values of other individuals; understand how personal values develop; and analyze the implications of decisions made on the basis of personal values.
Functioning Within Social Institutions:	Can identify those activities and institutions which constitute the social aspects of a culture (for example, governmental and economic systems, religion, marital and family institutions, employment, and civic volunteer and recreational organizations); understand the impact that social institutions have on individuals in a culture; and analyze one's own and others' personal functioning within social institutions.
Using Science and Technology:	Can identify those activities and products which constitute the scientific/technological aspects of a culture (for example, transportation, housing, energy, food, clothing, health maintenance, entertainment and recreation, mood alteration, national defense, communication, and data processing); understand the impact of such activities and products on the individuals and the physical environment in

a culture; and analyze the uses of technological products in a culture, including one's personal use of such products.

Using the Arts: Can identify those activities and products which constitute the artistic aspects of a culture (for example, graphic art, music, drama, literature, dance sculpture, film and architecture); understand the impact that art, in its various forms, has on individuals in a culture; and analyze uses of works of art within a culture and one's personal use of art.

These are assessed through the various reasoning, speaking and writing subtests described below. The test also yields scores in two derivative areas -- writing and speaking. Thus, the same performances appear to be scored for both knowledge of content and writing or speaking skill.

Description:

The COMP is designed for college students. There are alternative performance and objective tests, and an additional self-report of out-of-class activities that are related to the skills measured by the COMP. There are three secure forms.

The Composite Examination is a series of 15 simulation activities based on TV documentaries, recent magazine articles, ads, etc. Six of the simulations relate to assessing reasoning and communicating, three are writing samples, and three are speaking assignments. (The materials do not make clear what the other three activities consist of.)

Six of these simulations provide information on speaking -- three from the reasoning/communicating subtest and three from the speaking skills subtest. The three reasoning simulations require communicating about social institutions, science and technology and the arts. Written and audiotaped stimuli are used as a context for role-playing tasks in which participants speak to a friend, to an informal group, and at a formal meeting. Each task calls for endorsing a particular point of view and developing several specified points into a persuasive argument. As part of the reasoning and communication subtests, speaking is rated on the ability of the student to make and sustain contact with a relevant audience, organize a persuasive message that develops a number of relevant ideas, and present ideas clearly without hesitation and with energy and variety in voice quality. The six reasoning tasks (of which only three require speaking) take two hours to administer. The oral activities are usually taped in a language lab setting. It takes about 45 minutes per examinee to evaluate the responses.

The speaking skills assessment consists of three 3-minute speaking assignments (one-to-one, small group and large group) based on print stimulus materials that are usually given to students a day in advance. The entire assessment takes about 30 minutes to administer and 12 minutes per pupil to score. Administration is usually in a language lab setting with groups of students. Speeches are rated in the same manner as in the reasoning subtests -- appropriateness for the audience, quality of discourse (organization of ideas) and quality of delivery (vivid language, use of illustrations, etc.).

Scoring is done locally. Sample student performances and detailed scoring instructions are provided to users although they were not included in the materials we obtained.) ACT will rescore 10% of the writing and speaking samples to make sure that appropriate judgments were made.

Using our descriptive scheme, the COMP speaking assessments can be described as:

Purposes:	Transmitting information, analyzing messages
Setting:	One-to-one, small group, large group; formal language; one-way communication
Audience:	Peers, evaluators
Content:	Artificial, expository, persuasive
Responses:	Performance, skills in concert, impromptu and rehearsed
Level:	Communication competence

In terms of providing information that would enable one to select an instrument, the materials we received from COMP are rated as "fair" - "good". They discuss the populations recommended for use, the purposes of the instrument, the technical qualities of the test, and note administration and scoring requirements. They also provide samples of the tasks presented to students. They do not discuss development, the theoretical perspectives on which the test is based, use with special populations, or the limits of the test with respect to what it attempts to measure.

Reliability:

All reliabilities reported below relate only to the speaking scores.

Interrater reliabilities are "good" to "excellent." They range from .87 to .99, with a number of studies reporting reliabilities above .95.

Parallel form reliabilities from three studies range from .75 to .84. These are "fair" to "good."

Internal consistency reliabilities from several studies range from .88 to .92. These are "good."

Validity:

Several lines of research have taken place with respect to the COMP. (Many of the results reported below relate to the total score from the COMP and not to the individual speaking scores.)

1. Five studies are reported in which scores on the COMP are compared to supervisor ratings for various groups of adults in a number of employment settings (volunteers, bank employees, business/criminal justice management, practice teachers, student nurses). Generally, the overall COMP score was moderately related to composites of supervisor ratings. The speaking scales tended to have lower correlations with supervisor ratings than other scales in the assessment. These ranged from .16 to .34.

2. One study of 174 college graduates related COMP scores to an index of adult functioning based on occupational prestige, amount of volunteer activity and education beyond the baccalaureate degree. The relationships were moderate (.24 to .39). The relationship was about the same for the various ethnic groups in the sample.

3. Correlations between reasoning and speaking/writing are moderate (.37 to .52). Thus, these scales are somewhat related, but also measure some things that are independent.

4. Relationship between the COMP speaking score and other measures of achievement (GPA, ACT and a reading test) are low to moderate (.14 to .37), showing that it does not simply reflect differing levels of academic achievement.

5. The instruments appear to differentiate between college freshmen and seniors. This is due to the effects of a college education and not due to age maturation because a separate study of the scores of various age groups showed few differences in performance.

Based on these studies, we rate the COMP as "good" in terms of validity.

Help With Interpretation:

Norms appear to be based on users. Thus, they are not necessarily nationally representative. However, the norms are based on a number of different institutions, and represent a large number of students (1600 to 4000) depending on the age (freshmen or seniors) and subtest.

A criterion-referenced standard for performance is also suggested. This is the middle level of performance as defined by the rating scales. Thus, performance can be stated as the percentage of students that achieve this middle level of functioning. There is no rationale provided for this standard.

ACT also provides on-site consultations in assessment, education program development and improving education.

No other assistance with interpretation and use of results (or information about what assistance in this area is available) is provided in the materials we obtained. This area is rated as "fair."

Comments:

This set of instruments has very good face validity and reasonably good validity shown through a number of studies. Materials sent to potential users could be improved in the amount of information supplied so that users can determine exactly what is assessed and how it can be used.

No reviews were found in Hammill, et al. (1989), Keyser and Sweetland (1987) or Buros Mental Measurement Yearbook (Mitchell, 1985; Conoley and Kramer, 1989).

Title:

Diagnostic Achievement Battery (DAB), 1984

Author(s):

Phyllis L. Newcomer and Dolores Curtis

Source:

Slosson Educational Publications, P.O. Box 280, East Aurora, New York 14052. Also PRO-ED, 8700 Shoal Creek Blvd., Austin, Texas 78758.

Author's Description Of Purpose:

"The DAB is a reliable, valid, and nationally standardized individual achievement test that can be used to assess children's ability in listening, speaking, reading, writing and mathematics." (Manual, p.1)

"The DAB is intended to accomplish four purposes: (1) to identify those students who are significantly below their peers...and who, as a result, may profit from supplemental or remedial help; (2) to determine the particular kinds of component strengths and weaknesses that individual students possess; (3) to document students' progress in specific areas as a consequence of special intervention programs; and (4) to serve as a measurement device in research studies..." (Manual p. 3)

Author's Description Of Subtests:

(Note: Only the listening and speaking subtests are described here.)

Story Comprehension-(SC):	The examiner reads aloud brief stories and asks the student to answer certain questions about them. The items start with a two-sentence statement requiring the student to answer only one question and progress in difficulty to lengthier paragraphs requiring students to answer five questions. In order to succeed at this task, the student must listen to and comprehend the story being read.
Characteristics (CH):	This subtest requires students to listen to a brief statement and to decide whether the statement is true or false...The child must interpret each sentence using knowledge of the characteristics of objects or events and the cognitive categories to which they belong. For example, "All trees are oaks."
Synonyms (SY):	The examiner says a word and the child must supply a word that has the same meaning. This format requires both receptive and expressive abilities.
Grammatic Completion (GC):	This subtest measures the ability to understand and use certain common morphological forms in English. The format requires the examiner to read unfinished sentences and the student to supply the missing morphological form. Among the items included are those that require knowledge of plurals, possessives, verb tenses, comparative and superlative adjectives, and so forth. For example, "Here is one tree. There are two _____."

Description:

The DAB is intended for use with students aged 6.0 to 14.11. Since the test is individually administered, it is paced by the teacher. There is one level and one form but there is a different starting point for students aged 6-8 and students aged 9 and above. There are 122 items on the listening and speaking subtests; not necessarily all items are given to each student. All items and stimulus materials are read by the teacher. The student is required to provide short, oral answers. There are no multiple-choice questions.

The Story Comprehension subtest consists of the teacher reading narrative and expository passages of increasing difficulty followed by one to five questions that require recall of facts, recall of sequence, inferring the feelings of a character, identifying the main idea, interpreting figures of speech, and defining vocabulary. The students cannot take notes. Therefore, there is a moderate memory load required by the test.

Teachers score answers right or wrong as they are given. There are only short answers, and there is little interpretation required as to the adequacy of a response.

The instrument can be characterized as:

Purpose:	Transmitting information
Setting:	One-to-one, one-way communication, formal language, classroom
Audience:	Teacher
Content:	Artificial, narrative and expository passages
Response:	Short answer, skills in isolation, impromptu
Level:	Linguistic and communication competence

The information needed for a user to select the test is rated as "fair" - "good." There is a general lack of description of the theoretical basis for what the listening and speaking subtests are trying to accomplish, and therefore what inferences can really be made about the results.

Reliability:

The authors provide both internal consistency (coefficient alpha) and test-retest reliabilities. Overall, composite listening and speaking internal consistency reliabilities are good (medians across grade levels of .90 and .88, respectively). Some combinations of subtests and grade levels have substantially lower reliabilities; subtest reliabilities range from "poor" to "good."

Test-retest reliability is "good" - "excellent," but it is based on a very small sample.

Validity:

Validity studies included:

1. Content was selected to reflect commonly used curriculum and teacher programs.

2. Items were pilot-tested; item statistics are available.

3. Assessment formats were developed to match the requirements of each domain and, at the same time, be easy to use. Formats were reviewed by measurement experts to verify these considerations.

4. Correlations with other, related measures were provided. Each subtest was correlated with one other test that was identified as measuring the same content. These correlations (except for SY) were moderate. The correlation for SY was not statistically significant.

5. Scores increase as grades increase.

6. All subtests are highly interrelated. The authors predicted this because all the subtests related to communication.

7. Correlations with ability measures are moderate. This was expected because the communication skills on the test require cognitive processes.

8. There were significant differences in performance between a normal and a learning disabled population.

In general, the listening and speaking subtests of this test have been examined in more detail than those in other achievement test series. This test is probably a reasonably good measure of linguistic competence. However, the information presented does not answer the question of whether performance on this test is an adequate reflection of the daily performance of students in typical learning situations. Also, the test does not measure communication competence except in the area of listening comprehension. The number of students involved in many of the studies is very low. Therefore, we rate validity as "fair."

Help With Interpretation:

Norms for 12 age groupings are available. However, since only about 1500 students were tested, this means that norms are based on only about 125 students per grade.

There is other help with interpretation. Cautions with respect to the use of the test are given -- the test is only one piece of information; remediation should not be planned around the subtests because they are only a sample from the communication domain. There are some suggestions for expanding the test to get at student cognitive processes, motivation, etc. The authors provide some references for further assessment and instruction.

The rating of help with interpretation is "fair" to "good."

Comments:

The test received favorable reviews by Hammill, et al. (1989) and Keyser and Sweetland (1985) except with respect to norms. One review in Buros Mental Measurement Yearbook (Mitchell, 1985, 9:333) was also very positive.

There has indeed been more of an attempt to look at validity with this test than with other achievement test batteries. However, from the perspective of this Guide, the test is limited because it is more a measure of linguistic competence than communication competence. The range of skills assessed is very limited with respect to contexts, purposes and skills.

Oral, open-ended responses are advantageous because they minimize the need to read and require production rather than identification of the right answer. However, there is some memory load required on the listening comprehension subtest. A few questions also appear to require general knowledge. Reliabilities are good.

With respect to the authors' purposes, the test appears adequate for screening. However, I would question its use to determine student strengths and weaknesses or to document progress except as it relates to the limited areas covered by the test.

Title:

The English Language Skills Profile (TELS), 1987

Author(s):

Carolyn Hutchinson, Alastair Pollitt and Lillian Munro, University of Edinburgh

Source:

MacMillan Education, Houndmills, Basingstoke, Hampshire RG21 2XS, Great Britain

Author's Description Of Purpose(s):

"TELS Profile...is designed both to develop and to measure pupils' competence in language using a 'total language' approach which seeks to foster in children a broad range of language skills." (Manual, p. 8)

"TELS Profile is designed to be used in the classroom, by teachers and pupils...and, wherever possible, it is suggested that pupils be involved in the assessment of the exercises in the TELS Profile package." (Manual, p. 10)

"...it will help both pupils and teachers to identify areas of weakness in pupil's performances, and to plan for their remediation." (Manual, p. 18).

Author's Description Of Subtests:

(The entire test covers study skills, reading, listening and oral communication. We will only discuss the subtests on listening and oral communication.)

Productive Skills: The tests in this section are designed to measure how well pupils construct and produce spoken text, taking account of the purpose and audience for whom they are speaking. The group discussion "is designed to stretch the imaginative powers of the pupils by involving them in devising a group strategy to cope with an unusual set of circumstances." The purpose of the exercise is to assess each pupil's contribution to the group's discussion and the operation of the group as a whole. The purpose of the paired interview is to assess the ability of pupils to engage in different types of talk ranging from describing, explaining and analyzing to evaluating alternatives, seeking information and synthesizing in order to reach a conclusion at the end of the interview.

Description:

The test was developed for secondary level students -- grades 7 and above. There are two levels and two forms. Selection of level is based on student ability, not grade level. The two forms are not strictly parallel -- one emphasizes a theme of relationships and the other emphasizes a theme of community. However, the same subtests and general skills are covered by each.

The Listening test consists of listening to three passages originally broadcast over the radio -- narrative, personal experiences, and persuasive. Students are asked to answer questions requiring recall of details, summaries, inferences and speaker's style. All passages are on tape. Good features are that students are told what to listen for before the tape is played and students are encouraged to take notes while the passage is played. After the passage, students read and

answer questions in their test booklet. These are cloze, multiple-choice and short answer. There are around 40 questions and the test takes about 45 minutes to complete. A tape-recorder is required. It is recommended that the test be given in groups no larger than 15.

The group discussion consists of having a group of 4-5 students devise a strategy for coping with a presented emergency. Students read the instructions on Task Cards and have 15 minutes to come to a decision. The discussion is taped. Students analyze the tape themselves. They rate each contribution as to type (e.g., proposing, building, clarifying, reacting, and controlling) and quality (e.g., incomplete, ineffective). A tape-recorder is required. Students must be able to read and understand the Task Cards.

In the paired interview, pupils are given written information about a proposed project, and are asked to discuss in pairs various aspects of its implementation with a view to making decisions. There is an adult "interlocutor" at the interview. The students can ask questions of the interlocutor if they feel they need additional information. Performance is rated by the teacher on a five by five matrix (skills by discourse mode). Skills are: appropriateness (of register, accent, idiom and behavior); coherent fluency (in organization and sequence of ideas); superficial fluency (of speaking); interactive skills (when to take a turn, being able to sustain a point of view, ability to cope with disagreement, etc.); and amount of support (how much help the student needs to complete the task). Discourse modes are: describing, explaining, analyzing, evaluating, and seeking information. Thus, students can evidence each skill while engaging in the various discourse modes required for the task. The discussion is taped. There is no estimate of the time required for the interview or the scoring. Although the scoring rubrics are described in detail, there are no sample student "anchor responses" provided. This procedure would require training.

According to our descriptive framework, the instrument can be described as:

Purposes:	Transmitting information, analyzing and evaluating messages
Setting:	Small group, one-to-one, formal and informal language, interactive communication and one-way communication
Audience:	Teacher, peers, other adults
Responses:	Multiple-choice, short answer, performance; skills in concert
Level:	Communication competence

We rate the manual as "good" in terms of the information provided to assist with selection. The instrument is clear on the theoretical basis of the tasks and their limitations.

Reliability:

Internal consistency reliability for the listening subtest is .83 ("good") and for the oral communication subtests .94 ("excellent"). The latter is based on only a small sample size. There are no estimates of inter-rater reliability currently available for the pair-interview task.

Validity:

Validity considerations included:

1. There is strong theoretical background presented for the philosophy of the test as a whole and for each individual subtest.

2. All subtests were extensively pilot-tested and revised as the result of the piloting. Some features of the final tests are the result of the piloting -- for example, self-evaluation in the group discussion (teachers could not identify speakers), and reading multiple-choice

questions on the listening test rather than having them dictated (students were bored by the taped presentation).

3. IRT procedures were used to generate item statistics and to select items.

4. Ecological validity was addressed by seeing how well teachers and students could use results.

No other validity studies are provided at this time. The rating is "fair."

Help With Interpretation:

There are no norms available. However, there is extensive help with interpretation and use of results including profiling (using standard scores), discussions with students and planning instruction. More help could be given on how to score the interview task. The rating is "good."

Comments:

This instrument has very good face validity and attempts to directly address communication competence as defined in this Guide. However, there is still work to be done on validity, especially how reading ability interferes with performance, how well performance relates to daily communication skill (because of the artificiality of some of the exercises), how general social skills affect performance, and interrater reliabilities.

We found no reviews in Buros Mental Measurement Yearbook (Conoley and Kramer, 1989; Conoley, et al., 1988), Keyser and Sweetland (1987) or Hammill, et al. (1989).

Title:

Profile Of Nonverbal Sensitivity (PONS), 1979

Authors:

Robert Rosenthal, Judith A. Hall, M. Robin DiMatteo, Peter L. Rogers and Dane Archer

Source:

Irvington Publishers, 551 Fifth Ave., New York, New York 10017, (212) 777-4100.

(Note: The description of the instrument provided below is based on information in the book *Sensitivity to Nonverbal Communication – The PONS Test*, Baltimore: Johns Hopkins University Press, 1979, by the authors listed above. We were not aware of another source of this instrument until just prior to publication, and were not able to obtain a copy of the published version in time for this review. However, the information presented below has been reviewed by the authors for accuracy.)

Author's Description Of Purpose:

The purpose of the PONS is to measure the nonverbal decoding abilities of individuals and groups.

Description:

This test was designed for use with adults, but has been used with students down to grade 3. The test takes about 45 minutes and consists of 220 two-second segments of nonverbal behavior presented on videotape. Twenty different interpersonal situations are presented, each appearing 11 times with different combinations of face, body and tonal cues. The examinee must choose the situations being portrayed. All items are in multiple-choice format. There is one form. The same test is used for all age groups; the only difference being simplified answer choices for children.

Using our descriptive scheme, this instrument can be described as:

Purpose:	Transmitting information
Setting:	One-to-one, one-way communication
Content:	Artificial
Audience:	Assessor
Responses:	Multiple-choice, skills in isolation
Level:	Communication competence

We were unable to rate the manual on how well it provides the information necessary for selecting and using the instrument because we did not have the manual for review.

Reliability:

Internal consistency reliability of the total score is .86. This is "good." Reliabilities of channel scores are lower than for the total test score and are "fair" to "excellent" depending on the channel. Test-retest reliability averages .69. This is "poor."

Validity:

A great deal of information is available on the PONS. This includes factor analyses; effect of the length of exposure of the stimuli; cultural variation; other cognitive, affective and performance correlates; performance differences with age and gender; comparisons of impaired and normal groups; comparisons of people in different occupations; comparison of scores with supervisor ratings, etc. Overall, the ability of the PONS to measure nonverbal communication is rated as "good" to "excellent."

Help With Interpretation And Use:

We were unable to rate this area because we do not have the actual manual that is provided with the assessment materials.

Comments:

Rubin and Mead (1984) agree that the "test stimulus appears to have high ecological validity for the range of nonverbal sensitivity measured" (p. 90). There may be some confounding of nonverbal skills by ability to read and knowledge of the behavioral terms used.

Title:

Watson-Barker High School Listening Test (HS-WBLT), 1989

Authors:

Kittie W. Watson, Larry L. Barker, and Charles V. Roberts

Source:

Spectra, Inc., P.O. Box 1708, Auburn, Alabama 36831-1708.

Authors' Description of Purposes:

"The high school version measures the listening abilities of high school students -- grades 7 through 12" (Facilitator's Guide, p. 1). The authors' recommended uses include student self-awareness of how their listening skills compare to those of other students, administration as an instructional technique, pre- and post-testing to measure student growth, curriculum evaluation, identifying skills that need improvement and use in research.

Authors' Description of Subtests:

The test has five parts: interpreting message content/short term memory, understanding meaning in conversation, remembering lecture information/long term memory, interpreting emotional meaning and ability to follow instructions/directions.

Description:

The High School Watson-Barker is an adaptation of the adult version of the Watson-Barker for use in grades 7-12. There is one level and two forms. Each form has five subtests containing a total of 50 items. The test takes about 35 minutes to give and is administered using either a videotape or an audiotape. All instructions, pacing, passages and questions are incorporated into the tapes. Thus, the test is very easy to administer. Answer sheets do not reproduce the questions asked. All items are multiple-choice.

(Note: The content description below was derived from both the manual and examining the items.)

The five subtests consist of: (1) sentence comprehension (a sentence is read and students have to identify another sentence closest in meaning or best supported by the first sentence); (2) understanding social conversations (students hear seven conversations and answer one to three questions about each; most questions require literal comprehension of what was said); (3) understanding short expository and functional passages (five questions on each of two passages; questions that mainly require recall of facts); (4) interpreting other verbal and nonverbal cues (students identify the meaning of a sentence by *how* it is said); (5) understanding instructions (three to four questions about each of three passages; most questions require factual recall). The test requires a moderate memory load. Students are not allowed to take notes or ask questions.

The recorded listening situations were designed to be representative of high school and home life settings. They include a variety of contexts, accents, sound levels, speech rates and video sound quality. A variety of situations is emphasized because different listening situations require different listening strategies. The listening situations are not designed to be highly involving and interesting because they are designed to reflect real life. The authors attempted to restrict the vocabulary level to grade 9.

Using our descriptive scheme, this instrument can be characterized as:

Purposes:	Unclear. Form the student's perspective, the purpose is probably exchange of information. From the test developer's perspective, the purpose might be the implied purposes in the individual passages.
Setting:	Unclear. From the students' perspective, the purpose might be one-to-one, one-way communication, formal language. From the developer's perspective, the purpose might be that implied by the passage.
Audience:	Unclear. From the students' perspective, it might be the teacher. From the developer's perspective, it might be an audience implied by the passage and question.
Content:	Artificial; narrative, expository and functional passages; home and school situations
Responses:	Multiple-choice, impromptu, skills in isolation
Level:	Linguistic and communication competence

We rate the materials "fair" in terms of providing the information needed to select or use the instrument. The manual includes some descriptions of content and complete transcripts of the passages and questions, but little information on the theoretical basis of the instrument, technical information, cautions, or definitions of terms.

Reliability:

Only alternate form reliability is provided. This is based on about 400 students in grades 7-12. The reliability for the total score is .53; subtests range from .11-.38. This is "poor." One reason might be that many, unidentified, extraneous factors are affecting test scores. The two forms are of unequal difficulties and have been equated only at the mean.

Validity:

Validity information includes:

1. The test was adapted from the adult version of the Watson-Barker.

2. Preliminary scripts were examined by high school teachers and students.

3. A statement in the manual says that "test scores have been subjected to relational validity tests, item analyses, reliability tests and descriptive analyses." No actual data is provided.

4. An independent study (Karr and Vogelsang, 1988) revealed a factor structure that supports the five dimensions of the test and shows that scores increase after instruction.

Evidence of validity is rated "poor" to "fair."

Help With Interpretation

Help with interpretation and use includes:

1. Average total and subtest scores for male and female junior and senior high school students. This is based on a fairly good sample size of 400 students. No indication of the sample characteristics is given.

2. There is a scale for converting numerical scores to verbal ratings ranging from "very poor" to "excellent." No rationale is provided for how the conversion ranges were determined.

3. Appropriate cautions about overinterpretation of scores are provided.

4. Information is provided on how to respond to student concerns about the test.

5. Instructional sources are provided, but these are not tied to test scores.

6. There is a plan to provide yearly user norms.

We rate the instrument "fair" in this area.

Comments

This is an interesting instrument because of the targeted age range and because of the videotape format. However, there is a lack of technical information provided with the materials.

We found no reviews of this instrument in Keyser and Sweetland (1987) or Hammill, et al. (1989). Two reviews in Buros Mental Measurement Yearbook (Conoley and Kramer, 1989, 10:384) praised the instrument for the quality of the tapes, but agree that evidence of validity and reliability is lacking.

APPENDIX C

Short Reviews

SHORT REVIEWS -- RESEARCH INSTRUMENTS

This section contains reviews of instruments that were designed primarily for use in research rather than use in the schools. They all have some technical information provided, but because they were designed for research purposes they generally do not provide enough information (in the source listed) for using the instrument or interpreting results. For example, the source might only reproduce part of the instrument, or there is not enough information about the scoring procedure. This information must be obtained from the author. In addition, many of the instruments only report the performances for the students in the research study, and there is rarely assistance with using results in the classroom. There are no reviews of these instruments in Buros Mental Measurement Yearbook (Buros, 1978; Mitchell, 1985; Conoley and Kramer, 1989; Conoley, et al., 1988), Hammill, et al. (1989) or Keyser and Sweetland (1987). Because of these factors, the instruments in this section should only be used by those knowledgeable in the area of assessing speaking and listening.

Class Apprehension About Participation Scale, 1987

M.R. Neer (1987), Communication Education, 36, 154-166.

The purpose of the Class Apprehension About Participation Scale is to identify the level of student anxiety about participating in classroom discussions and asking/answering questions in class. It was designed for college level students, but could be used at the high school level. Students indicate the degree to which 20 statements apply to them. There is one form and one level. There are no estimates of the time required to take the survey, but probably no longer than 10 minutes.

Internal consistency reliabilities for the two sections of the survey are .88 and .91. This is "good to excellent." There was a factor analysis in which all the items were found to be related to a unitary factor. Responses to the measure were related to other classroom behaviors and instructional preferences. Validity, as a measure of class apprehension, is "good." Summary statistics are provided for the students in the study.

There is a second section that asks students to identify those aspects of teaching style and classroom procedures that make them more and less anxious.

Both sections of the survey instrument are reproduced in the source listed above.

Since this instrument looks at the affective component of communication, rather than at communication competence, we will not categorize it as to purpose, task, etc.

Interactional Competency Checklist, 1978

J. Black (1978). Research in the Teaching of English, 13, 49-68.

This instrument was designed for use with students in grades K-3. There is one form and one level of a 16-item checklist to be used by teachers to assess the interactional competence of young children. Interactions are rated in the areas of ability to adapt to changes in the setting, appropriateness of nonverbal communication, and knowing how to carry on a conversation. This checklist is to be used to rate students as they participate in a sociodrama (a play session with a

theme). No time estimates for length of play session or rating requirements are given. Students can be videotaped.

No reliability information is provided.

The instrument is based on the view that naturalistic assessment of young children's language is a more valid procedure than published, standardized tests using artificial tasks. Evidence for validity include: (1) content based on a literature review of communication competence; (2) ratings based on a model of interactional competency; and (3) study results cited below. Validity is rated as "fair."

The instrument was used in a study of whether the evaluation of kindergarten children's oral language in an informal context of the natural classroom environment provides more comprehensive information about children's communicative competence than the ITBS or the CIRCUS. The results showed that the sociodrama was much better than the standardized tests for assessing communication competence, and equal or superior in terms of estimating linguistic competence.

The source above does not include criteria for rating performances nor specifics about the nature of the sociodrama. Additional information would need to be requested from the author.

Purposes:	Social interaction
Setting:	One-to-one, informal, interactive communication, classroom
Audience:	Peers
Content:	Naturalistic
Responses:	Performance, skills in concert, impromptu
Level:	Linguistic and communication competence

Language Communication Skills Task (LCST), 1972

M.C. Wang, S. Rose and J. Maxwell, Learning Research and Development Center, University of Pittsburgh, Pittsburgh, Pennsylvania 15213. Also in ETS Tests in Microfiche, ETS, Princeton, New Jersey.

The LCST was designed for students in grades K-2. There are two "referential communication" tasks in which two students sit across from each other, and one tells the other where in a picture to place various objects. One picture is of a classroom; the other is of a kitchen. The students alternate being the presenter and the receiver. Although the players are not permitted to look at each other's pictures, they can interact verbally as much as they want.

There is one form and one level. The tasks are untimed but take about 25 minutes for both. The verbal interaction is taped and scored in terms of both communication and linguistic competence. In the area of communication competence, the presenter is scored on correct labeling/description of objects and the correct description of placement of objects. The receiver is scored on the ability to select the correct object, place the object where it belongs and ask necessary clarifying questions. Linguistic competence is assessed by looking at the total number of words used, the total number of different words used, the average length of words, the average length of utterances and repetitiveness.

Internal consistency reliabilities are reported as .72 and .76. This is rated "fair."

A number of other analyzes were also performed to provide evidence on validity. This includes: (1) a high relationship between the various ratings of communication competence (e.g., correct labeling of objects) and successful placement of the objects; (2) significant performance

differences among children of different ages; (3) moderate correlations with achievement test results (this would probably be expected because the instruments would tend to measure different things); and (4) nonsignificant correlations with intelligence and gender. One interesting finding was that certain measures of linguistic competence were not related to the ability to successfully accomplish the task. Validity of assessing communication competence is "fair." There needs to be more study of how these tasks relate to everyday communication success.

Rubin and Mead (1984) conclude that "this test may provide useful data. However, more rigorous, systematic evaluation is needed before test users can be assured of adequate validity and reliability..." (p. 63).

The LCST is included in the short reviews because the entire instrument is not included in the references given, and would have to be requested from the author.

Purposes:	Transmitting information
Setting:	One-to-one, informal, interactive communication
Audience:	Peers
Content:	Artificial, descriptive
Responses:	Performance, skills in concert
Level:	Linguistic and communication competence

Notebook Communication Game, 1979.

W.P. Dickson, Center for Individualized Schooling, The University of Wisconsin, Madison, Wisconsin.

The Notebook Communication Game was designed to study referential communication performance -- how well one person can communicate a task to another person. The instrument has been used with children age 4-8 and with adults. The task is for one person to get another person to choose one of four pictures through description alone. Usually, each person in the pair has a chance to be both sender and receiver of information. The score is the number of errors made before the target picture is correctly identified. There is one form and one level. There are 12 items.

The instrument has been used in a number of studies, but the results are not reported in the source cited above. Further information about administration and use would have to be requested from the author.

Purposes:	Transmitting information
Setting:	One-to-one, informal, interactive communication
Audience:	Peer, parent, teacher
Content:	Artificial, description
Responses:	Multiple-choice, performance, skills in concert
Level:	Communication competence

Personal Report of Communication Apprehension (PRCA-24B), 1986.

J.C. McCroskey (1986). An Introduction to Rhetorical Communication, Englewood Cliffs, New Jersey: Prentice-Hall.

This is a short questionnaire designed to provide an indication of how much apprehension one feels in a variety of communication contexts. It was designed for college level students, but could be used at younger ages. There is one form and one level. There are 24 questions covering

anxiety about communication in four settings (talking at a meeting, interacting in a small group, conversing with one other person and public speaking) with three types of audiences (strangers, acquaintances and friends).

There is no technical information provided in the source listed above, although this source does reference earlier articles in which such information is presented. We were not able to review this additional information in time for publication.

One review (Leary, 1988) describes the internal consistency reliability for the total score to be above .90; subscales are above .85. This is "good." This same source describes a number of studies bearing on validity. He reports that "criterion validity is excellent," although construct validity information is still lacking. Because we were not able to review evidence ourselves, we will not rate the instrument on validity.

The instrument is not described using our system of purposes, settings, audiences, etc., because it is a measure in the affective domain.

Two Referential Communication Tasks, 1979.

W.P. Dickson, N. Miyake and T. Muto, Center for Individualized Schooling, The University of Wisconsin, Madison, Wisconsin.

This document presents two "referential communication" tasks designed for use in research at the college level. The tasks could also be used at the high school level. In one task, one student has three minutes to orally direct another on how to build a block structure. Students can interact verbally with each other. Students are scored on the number of blocks correctly placed. Since performance depends on another person, it is suggested that each person to be assessed be paired with a number of others in both the receiver and sender roles. The score is the total number of correctly placed blocks in all trials.

In the other task, the experimenter reads 64 different descriptions of 16 abstract pictures to the group as a whole. Students match the descriptions with the pictures. Students may not ask questions. Students are scored on how many they get right.

There is some technical information available, but it is restricted to overall performance and relationships between performance on the two tasks. Reliability and validity are rated as unknown.

Purposes:	Transfer of information
Setting:	One-to-one, informal, interactive and one-way communication
Audience:	Peers, teacher
Content:	Artificial, descriptive
Responses:	Performance, multiple-choice, skills in concert
Level:	Communication competence

Willingness To Communicate Scale (WTC), 1987.

J. C. McCroskey and V.P. Richmond (1987). Willingness to Communicate. In J.C. McCroskey and J.A. Daly (Eds.) Personality and Interpersonal Communication, Beverly Hills, CA: Sage.

The WPC was developed to measure the willingness of persons to communicate in various contexts (public speaking, talking in meetings, talking in small groups and talking in dyads) to various types of receivers (strangers, acquaintances and friends). There are 12 scored items and 8

filler items. Respondents indicate the length of time they would be willing to communicate to various receivers in various contexts. Subscores can be calculated for each context and receiver. The instrument appears to be developed for adults, but could probably be used in high school. There is one form and one level.

Internal consistency reliability is .92 for the total score and range from .65 to .82 for the subscores. These are "fair" to "good."

Validity information includes: (1) content based on previous research; (2) a factor analysis that shows that all items seem to measure a single factor; (3) moderate intercorrelations between the subscales; and (4) willingness to communicate decreases with the number of receivers and the distance of the relationship of the individual with the receiver. This evidence is rated "fair."

This instrument is not described on our general categories of task, purposes, etc. because it measures an affective area.

SHORT REVIEWS – ACHIEVEMENT TEST SERIES

(Note: Only the listening and speaking portions of achievement tests are reviewed.)

Most of the achievement test batteries we reviewed are included here as short reviews. Although they are readily accessible and have a listening subtest, they generally are not explicit in terms of the theoretical perspective of the listening test, and generally do not provide validity information explicitly for the listening subtest except general item statistics and content review.

The listening subtests in the achievement test batteries described below entail the teacher reading sentences/passages and multiple-choice questions to students. The tests usually cover some combination of linguistic and communication competence including receptive vocabulary, understanding sentences of various levels of syntactic and grammatical complexity, auditory memory, and answering recall and inference questions about passages. None of the achievement test batteries described here have speaking subtests. These tests can generally be characterized by:

Purpose:	Unclear. From the students' perspective, the purpose is probably exchange of information. From the test developer's perspective, the purpose might be the implied purposes in the individual passages.
Setting:	Unclear. From the students' perspective, the purpose might be one-to-one, formal language, one-way communication. From the developer's perspective, the purpose might be that implied by the passage.
Audience:	Unclear. From the students' perspective, it might be the teacher. From the developer's perspective, it might be an audience implied by the passage and question.
Content:	Narrative passages at the lower levels. Persuasive and expository passages are sometimes added at the higher levels. All tasks are artificial as opposed to naturalistic.
Responses:	Multiple-choice, skills in isolation, impromptu
Level:	Linguistic and communication competence

Thus, achievement test series are somewhat limited in terms of the purposes, contexts, skills, content and responses that would sample from the entire domain of communication competence or even "oral language skill." Even though many of the instruments have good face validity for listening comprehension, rigorous general review and standard item statistics, their use and interpretation is somewhat limited because of their lack of specialized validity studies and lack of explicitness in terms of the theoretical underpinnings for the content.

The instruments differ in terms of:

1. Their relative emphasis on linguistic or communication competence. For listening comprehension tests it is often hard to distinguish these. We use the term linguistic competence when the major tasks are vocabulary, literal understanding of phrases and sentences of various levels of complexity, grammar, ability to use different descriptive categories, etc. We use the term communication competence when the test requires listening to passages and answering questions requiring factual recall and inferences.

2. The specific skills covered. Some emphasize more recall of facts and some emphasize more inference.

3. The types of listening passages -- narrative, expository, persuasive, and/or functional; also the attempt to supply "real-life" material.

4. Whether teachers read the question to be answered before or after the passage itself.

5. The grade levels covered by the listening subtest.

In general, *with respect to the listening components* of the tests, the instruments can be rated as "fair" - "good" in terms of the information presented to the user to enable them to select an instrument, and "fair" - "good" on assistance with interpretation and use. Ratings would be higher if the tests were more explicit about the theoretical underpinnings of the items, and provided more validity information. All of the tests have good norms. Individual ratings on reliability and validity will be given as part of the reviews below. No reviews from other sources will be included unless they deal specifically with the listening portions of the tests.

California Achievement Test (CAT), 1985

CTB/McGraw-Hill, 2500 Garden Road, Monterey, California 93940, (800) 538-9547.

The CAT is an 11-level achievement test battery covering grades K-12. There are two forms for each level. At Level 10 (Grade K) the reading subtests resemble the listening vocabulary and comprehension subtests of other test batteries. The vocabulary subtest (30 questions) requires students to pick the picture of a word that is read, or to find the picture of a word that has been left out of a sentence (cloze format).

The comprehension subtest (22 questions) requires students to match a picture with a sentence and to pick a picture that answers a recall, inference or main idea question about a short, narrative passage. Students are told to listen carefully as the story is read aloud to them and then are asked questions about the story. For the items based on single sentences, the emphasis is on linguistic competence. For the items based on a passage, communication competence is emphasized. Working time appears to be 90 minutes.

Information on reliability was not included with the sample materials.

No rating on validity is given because the test was intended to measure a prereading skill and not listening.

There is a supplemental listening test (see Listening Test below).

Comprehensive Tests Of Basic Skills (CTBS), 1989

CTB/McGraw-Hill, 2500 Garden Road, Monterey, California, 93940, (800) 538-9547.

The CTBS is an 11-level achievement test battery covering grades K-12. There are two forms for each level. The reading subtests at grades K-2.2 (Levels K, 10 and 11) have some portions that correspond to those called listening vocabulary and listening comprehension in other achievement test series. The intermediate level (Level 10) was specifically designed to serve as a transitional link between oral and written communication.

The vocabulary subtest has two parts -- cloze, in which students choose the picture of the word that is missing; and direct, in which students identify the picture of a word that is read. At Levels 10 and 11 the subtest also entails finding the written word of a word that is orally defined. One cloze

item type specific to Level 10 combines oral and reading comprehension by asking students to read a "short story" (one or two sentences) while the teacher reads the story aloud. Then the students choose the written word that best fits in the missing part of the story.

The comprehension subtests involve picking a picture that illustrates a sentence; or picking the picture that answers a recall, inference or main idea question about a short narrative passage. Some of these questions require students to make predictions and to differentiate between reality and fantasy. The authors feel that "these and other inference questions demonstrate a greater communication emphasis than is usually found in listening tests, requiring the application of higher level thinking skills to the comprehension of orally communicated information. This represents a planned approach to the listening component based on an integrated view of language arts."

One to three questions are asked about each passage. There is thus some memory load. At Levels 10 and 11, students are also required to read and understand sentences as part of the comprehension subtest. At Level 10 the passages are read in short parts with questions on that part immediately following, except for a few general questions at the end of the passage. The emphasis is on both linguistic and communication competence.

For the vocabulary and comprehension subtests, there are 48, 60 and 66 items (for levels K, 10 and 11, respectively) taking about 38, 48 and 55 minutes to give. Internal consistency reliabilities range from .72 to .89. Thus reliability is "fair" to "good" depending on level and subtest.

The test is not rated on validity because the original intent was to measure prereading skills, not listening comprehension.

There is a supplemental listening test that ties in with the achievement battery (see Listening Test).

Comprehensive Testing Program (CPT II), 1982

Educational Records Bureau, Bardwell Hall, 37 Cameron Street, Wellesley, Massachusetts 02181, (617) 235-8920.

The CTP-II is a five-level achievement test battery covering grades 1-9. It is published by the Educational Records Bureau which requires membership in order to purchase its materials. Their tests are designed to measure the best kids; ERB says that the CPT-II has a higher ceiling than other test series. A listening subtest is included at Levels 1 and 2 (Grades 1-3). There is only one form of this subtest, although other subtests in the battery have two forms.

The listening subtest assesses children's ability to comprehend words, sentences or paragraphs, and recall, interpret, evaluate and draw inferences about sentences and paragraphs. One to three questions are read after each selection. The test covers both linguistic and communication competence. The listening subtest has 40 items and takes about 40 to 60 minutes to give.

Internal consistency reliabilities of the listening subtest range from .66 to .78 depending on level. This is "fair."

Validity considerations include: (1) the tests were developed to match the curricula of member schools, including review of content by teachers, and (2) correlations between the listening subtest and other subtests are moderate (this would be expected). This is rated as "poor" - "fair."

Norms are based on equating the tests to the CIRCUS/STEP. Thus, no empirically derived norms are available for the CTP-II. In addition, the CIRCUS/STEP norms are very old (1976-77).

Iowa Test Of Basic Skills (ITBS), 1990

Riverside Publishing Company, 8420 Bryn Mawr Ave., Chicago, Illinois 60631, (800)323-9540.

The ITBS is a 10-level achievement test battery covering grades K-9. The upward extension is the Tests of Achievement and Proficiency. Listening subtests (two forms) are included as part of the battery at grades K-3.5 (Levels 5-8). Listening tests (one form) can be obtained as a supplement to the battery at grades 3-8 (Levels 9-14).

The listening subtest in grades K-3.5 requires picking a picture that illustrates a sentence or answers a question about a short narrative passage. At levels 5 and 6 specific skills covered by the test are literal meaning, inferential meaning, concept development, following directions, understanding sequence, predicting outcomes and attention span. Additional skills at levels 7 and 8 are linguistic relationships and numerical and spatial relationships. Questions are read after the passages. Both linguistic and communication competence are addressed. The tests are teacher-paced, but take about 25 (Levels 5 and 6) or 16 minutes (Levels 7 and 8) to give. There are 31 items on Levels 5 and 6, and 32 items on Levels 7 and 8.

Internal consistency reliabilities for the listening subtests range from .64 to .78 (median .72) depending on level and time of year. These are "fair."

Information on validity includes: (1) content validity based on curriculum review, expert opinion and interaction with users; (2) moderate predictions of later teacher ratings of reading and reading readiness; (3) high correlations between listening and the other subtests on the ITBS, indicating that they all measure common aspects of achievement (as expected); (4) a factor analysis (determining the underlying structure of the test) in which listening skills did not load with any other skills, indicating that some aspects of this subtest are unique (a desirable state of affairs). Validity is rated as "fair.". There could be more information on how performance on the test relates to actual success in the classroom in terms of communicating for various purposes.

The listening supplement for grades 3-8 has 95 items in a multi-level booklet (6 levels) -- students at different grade levels begin at different item numbers. Teachers read short narrative, expository, persuasive or functional (e.g., report of a crime) passages, followed by three to ten questions requiring the student to recall details, make inferences, follow directions, identify the speaker's purpose, point of view or style, and define words. There is a heavy memory load on this portion of the test. There are also a few short questions not relating to any passage that require mental arithmetic, number sequences, etc. These questions cover both linguistic and communication competence. There is one form per level.

Reliabilities range from .70 to .81 depending on grade. These are "fair" to "good." Validity information comes from the same source as that reported for the ITBS general battery. Validity is again rated as "fair" for the same reason as above.

Language Diagnostics Test, 1988

Psychological Corporation, 555 Academic Court, San Antonio, Texas 78204, (800) 228-0752.

The Language Diagnostics Test is a 9-level achievement test battery designed to complement the MAT-6 survey tests (see below). It covers grades 1-9. There is one form for each level. Listening comprehension is included as a subtest for Levels P1-E (Grades 1.0-4.9). (The publisher states that it is not included for higher grade levels because they found that most students in higher grades already possessed the skills covered.)

In the listening test the teacher reads the stimulus materials (one to several sentences) and the student chooses a picture that answers a question. Questions mainly cover linguistic competence -- matching a picture to a description, rhyming words, syntax, pronouns/referents, and negatives. Some questions require the students to listen to a short narrative passage and answer a question which requires recall of facts, main idea and sequence of events. All questions are read to the students before the sentence or passage. There are 192 questions.

Reliabilities range from .63-.68. These are "poor" to "fair." Validity information includes: (1) content chosen to be reflective of current curriculum, (2) content review by experts, (3) moderate to high correlations with the MAT-6, (4) increased performance with grade level, and (5) measures of independence of the subtests. Validity is rated as "fair" as a measure of communication competence.

Listening Test, 1985

CTB/McGraw-Hill, 2500 Garden Road, Monterey, California, 93940, (800) 538-9547.

The Listening Test is a six-level battery covering grades 3-12. It may be used as an optional listening supplement to the CAT and the CTBS. There is one form for each level. The purpose is to "measure the ability to follow directions and interpret connected discourse."

Stimuli for the items are contained on a worksheet. For the "following directions" items, the examiner reads directions for a task and the students follow the directions in a work area to arrive at their answer. For example, "Start at the letter B. Go to the X and then to the A. Follow a straight line from A past X. At which letter do you end?"

The listening comprehension portion of the test entails listening to narrative and expository selections read by the teacher and answering one to five questions read aloud. Only answer choices are printed in the test booklet. There is, thus, somewhat of a memory load on the test. Skills include recall of information, sequence, main idea, knowledge of vocabulary and inferences. Communication competence is emphasized more than linguistic competence.

The test is not timed, but usually takes 30-40. There are 18 to 20 questions depending on level. The content was brought up to date in 1985 but the norms are based on those originally developed in 1979. No reliabilities or other technical information were provided with the samples we received.

Metropolitan Achievement Test, (MAT-6), 1987

Psychological Corporation, 555 Academic Court, San Antonio, Texas 78204, (800)228-0752.

The MAT-6 is an 8-level achievement test battery covering grades K-12. There are two forms for each level. Level PP (Grade K) has a language subtest of 24 questions in which students match a picture with a dictated sentence (linguistic competence only). There are some (fewer than 10) listening comprehension items included in the language subtests for Levels P-P2 (Grades K.5-3.9). These again involve matching a picture to a sentence that is read and thus emphasize linguistic competence. Testing time for the language subtest at Levels PP through P2 is 18 to 25 minutes.

No technical information was included in the samples we obtained.

There is an associated Language Diagnostics Test (see above) that covers listening comprehension more fully.

Metropolitan Readiness Tests, (MRT), 1986

Psychological Corporation, 555 Academic Court, San Antonio, Texas 78204, (800) 228-0752.

The MRT is the lower extension of the MAT-6. It was designed to predict later achievement in reading and math. It has two levels covering grades K-1. There is one form for each level. Level 1 has subtests for auditory memory (picking out the picture that shows three or four items in the order mentioned by the teacher), and school language/listening (matching a picture to a sentence, and recalling facts or making inferences based on a short passage). Level 2 has separate subtests for school language (matching a picture to a sentence) and listening (recall of facts and making inferences based on a short narrative passage). Questions are presented to students after the passage is read. The instrument appears to assess linguistic competence more than communication competence.

There are 27 items in these areas at Level 1 and 18 items at Level 2. These tests take about 30 minutes to give at Level 1 and 15 minutes at Level 2.

For these subtests internal consistency reliabilities ranged from .66 to .80. Test-retest reliabilities range from .68 to .82. This is "fair" to "good." The reliabilities for Level 1 are better than for Level 2.

Information on validity includes: (1) content based on a review of the literature related to early school learning (but, this is not described in detail); (2) a low-moderate correlation between the language scores and later performance on the MAT-6 and SAT; and (3) moderate correlations between subtests (the manual does not explain whether this is good or bad). The validity rating is "fair." There needs to be further work on how language performance relates to actual classroom performance.

There is an associated "Early School Inventory -- Developmental" checklist that teachers can use in the classroom. There are 14 ratings in the areas of speaking and listening. These cover both linguistic and communication competence. There is no technical information.

National Achievement Test (NAT), 1989

American Testronics, P.O. Box 2270, Iowa City, Iowa 52244, (800) 553-0030.

The NAT is a 12-level achievement test battery covering grades K-12. There are two forms for each level. It is tied in with the company's other products (Assessment of Writing, School Attitude Measure and Developing Cognitive Abilities Test) to form the Comprehensive Assessment Program. There are listening tests for the first three levels (Grades K-1). The listening vocabulary portion requires students to categorize words or to identify the picture or written form of a definition presented orally. Level C also requires analogies. The listening comprehension portion requires students to listen to a short narrative selection and answer questions involving literal recall and inferences (predictions, figurative language and drawing conclusions). Students respond by indicating a picture or a written phrase. We do not have complete test booklets so it was impossible to tell how many questions are associated with each passage. Both linguistic and communication competence are covered.

Depending on level, there are 55-71 items requiring about 40-50 minutes to give. Technical information was not provided along with the samples we received.

National Test Of Basic Skills (NTBS), 1985

American Testronics, P.O. Box 2270, Iowa City, Iowa 52244, (800) 553-0030.

The NTBS is a 12-level achievement test battery covering grades K-12. There are two forms of each level. The purpose of the test is "the measurement of student learning in the basic skills and subject areas taught in our nation's schools."

Levels P, A and B (Grades PreK-1.5) have a listening comprehension subtest. At level P, the 9 items covering auditory comprehension appear to emphasize general knowledge rather than listening comprehension (e.g., "If you had a broken leg, what would you use to help you walk?")

At level A, the listening comprehension subtest (30 items) requires the student to match a sentence to a picture. At Level B (20 questions) students are read short narrative passages and are asked one factual recall question about each. Level B also has a receptive vocabulary section (20 items) that requires students to match pictures to words or select a word that matches a definition presented orally. Both levels A and B emphasize linguistic competence.

The subtests described above take 30 minutes to give at Level A and 35 minutes to give at Level B. There is no estimate of administration time at level P.

The internal consistency reliabilities for listening comprehension at Level B are .75 (fall) and .78 (spring); for vocabulary these are .85 (fall) and .87 (spring). These are "fair" to "good." No separate reliabilities are provided for subtests at Levels P and A.

Validity information includes: (1) content based on a review of the curriculum materials in use in the schools and on expert opinion; (2) high correlations with another achievement test battery; (3) moderate correlations of listening scores with the other subtests (indicating that they all measure some common aspect of achievement); (4) a factor-analysis (to determine the structure of the test) which confirmed that there is a large common aspect measured by the subtests, proposed as being "language" (thus, the question arises as to whether the listening subtests measure anything different); and (5) moderate correlations with teacher ratings of achievement in language. This evidence is rated as "fair" because there was no specific examination of how the listening scores relate to daily performance using a broader definition of listening comprehension.

Stanford Achievement Test (SAT), 1989

Psychological Corporation, 555 Academic Court, San Antonio, Texas 78204-2498, (800) 228-0752.

The SAT is an 8-level achievement test battery covering grades 1-9. The SESAT is the lower extension and the TASK is the high school extension. There are two forms for each grade.

There is a listening subtest for all eight levels which involves both listening vocabulary and listening comprehension. The listening vocabulary subtest requires students to respond to a stimulus word by choosing the appropriate printed response. The listening comprehension subtest requires listening to a variety of short passages (expository, functional, narrative, persuasive and descriptive) and answering from one to four questions about each. (Thus, there is a moderate memory load associated with the test.) Both passages and questions are read to the student. Answer choices, but not questions, are also provided in the student test booklets. Questions require recall of information, sequence of events, identifying the setting, plot or theme, and inferences. The vocabulary subtest emphasizes linguistic competence; the comprehension subtest emphasizes communication competence.

There are 45 items at each level requiring 30 minutes of testing time. No technical information was provided with the specimen set.

Stanford Early School Achievement Test (SESAT), 1988

Psychological Corporation, 555 Academic Court, San Antonio, Texas 78204-2498, (800) 228-0752.

The SESAT is the lower extension of the Stanford Achievement Test Series. It has two levels covering grades K and 1. There is one form for each level. The listening subtests consist of listening vocabulary and listening comprehension. In the listening vocabulary subtest, students mark the picture of the word that is read. Listening comprehension consists of recalling information from or making inferences about a narrative passage read by the teacher. A good feature is that students are told what information to listen for before the passage is read.

The test covers both linguistic and communication competence.

The test is teacher-paced, there are 45 items and the listening subtests take about 30 minutes to give. Technical information was not supplied with the specimen set.

Survey Of Basic Skills (SBS), 1985

Science Research Associates, Inc., 155 North Wacker Drive, Chicago, Illinois 60606.

The SBS is an 8-level achievement test battery covering grades K-12. There are two forms for each level. The purpose of the SBS is to "survey students' general academic achievement."

There is a listening comprehension subtest for grades K-1 (Levels 20-21). This involves listening to narrative passages of increasing levels of difficulty and answering one question about each. The questions require recall of information, sequence of activities, following directions, identifying cause and effect, predicting what will happen next, inferring information about a character, and main idea. The students are told what type of question they will be asked before the passage is read. Communication competence is emphasized more than linguistic competence.

The test is teacher-paced, has 22 (Level 20) or 23 items (Level 21), and takes about 20 minutes to give. Internal consistency reliabilities for these subtests range from .67 to .73. This is "fair." The SBS is somewhat different from the other achievement test series in that specific instructional activities in the area of listening are suggested.

Information on validity includes: (1) content based on a review of textbooks, curriculum guides, and professional journals, and advice from curriculum experts; (2) moderate to high correlations between the listening subtest and the other subtests, suggesting that a common set of skills is being assessed; and (3) a moderate relationship between scores on level 20 and level 21, indicating prediction over time. This is rated as "poor" - "fair" evidence of validity. More information about the relationship of performance on the test to performance in real life situations is needed.

Tests of Achievement and Proficiency (TAP) -- Listening Supplement, 1987

Riverside Publishing Company, 8420 Bryn Mawr Avenue, Chicago, Illinois 60631.

The "TAP--Listening" is a supplement to the TAP, a four-level achievement test series covering grades 9-12. (The TAP, in turn, is the upward extension of the ITBS). There is one form for each level. Students answer in a multi-level test booklet -- students start and end at different places

depending on grade level and/or functional ability. All passages, questions and answer choices are read by the teacher. There are six sections. Two require listening to expository passages that require recall of facts, making inferences, and identifying main ideas and details. One of these is a lengthy simulated lecture with 10 questions and the other is a shorter passage with 10 questions. Therefore, there is large memory load on these sections.

The other sections of the test do not require responses to passages, but require remembering sequences of letters and numbers, knowledge of vocabulary, identifying fact and opinion, and identifying language that indicates bias and prejudice. Thus, the test covers both linguistic and communication competence.

Around 50 items are given at each grade level, with a working time of 40 minutes. Internal consistency coefficients range from .82 to .85 depending on the grade level and time of year. These are "good." There are specific suggestions for improving instruction based on the results. No other technical information is provided, but we assume that it is similar to that provided for the ITBS Listening Supplement.

SHORT REVIEWS –
OTHER COMMUNICATION COMPETENCE RELATED INSTRUMENTS

This section provides short reviews of instruments measuring some aspect of communication competence that are easily accessible but do not come with any technical information. Many of these instruments were designed for informal use in the classroom.

Assessing Children's Speaking, Listening and Writing Skills. The Talking and Writing Series, K-12, 1983.

L. Reed, Dingle Associates, Washington, D.C. Also ERIC ED 233 380.

This article describes some considerations involved in doing classroom assessments and provides some sample assessment ideas. There is one instrument in the area of speaking and listening. The Group Self-Rating Scale is used by students to rate their own group presentations. Ten yes/no questions are grouped under planning the presentation and doing the presentation. There is no grade designation, but appears to be useful in grade five and above. There is no technical information. No sample student discussions are provided to illustrate rating, and the document does not include sample discussion topics.

Purposes:	Transmitting information, social interaction
Setting:	Small group, classroom, one-way and interactive communication
Audience:	Peers, teacher
Content:	Naturalistic
Responses:	Self-rate, self-rate
Level:	Communication competence

Diagnosis Of Group Membership, 1953

L. Crowell, Speech Teacher, 2, 26-32.

This is old but has been cited recently as a scale for rating group discussions. Development was based on a survey of criteria by which instructors in college courses on discussion rate participants. The scale could, however, be used at lower grade levels. There are five analytic ratings (sensitivity to other members, objectivity of contributions, worth of information presented, worth of thinking done, acceptance of full share of group responsibility) followed by a holistic rating of the group as a whole. Each area is rated on a score of one to five. The rating form can be used during any classroom discussion. No technical information is available. No sample student responses are provided to illustrate ratings and the document does not include sample discussion topics.

Purpose:	Transmitting information, analyzing messages
Setting:	Small group, informal language, interactive communication
Audience:	Peers, teacher
Content:	Naturalistic
Responses:	Performance, skills in concert
Level:	Communication competence

Evaluating Classroom Speaking, 1981

D.G. Bock and E.H. Bock, Evaluating Classroom Speaking, Speech Communication Association, Annandale, VA. Also ERIC ED 214 213.

This monograph discusses in detail how to do a classroom speaking assessment. Several sample informal rating forms and checklists are included. These include rating an introduction to a speech; informative and persuasive speeches (organization, language, material delivery, analysis and voice); technical and business speaking (audience analysis, organization, credibility, research, delivery, and overall presentation); an oral interpretation (introduction, material, eye contact, articulation, facial expression, poise, bodily action, vocal quality, rate, content); and rating a group project (organization, participation, quality and creativity).

These various rating forms are ungraded but look appropriate for grades 5 through adult. There are 10 different rating forms provided, none of which is accompanied by technical information. No student responses are provided to illustrate the scoring and the document does not include sample topics for speeches.

Purposes:	Transmitting information, self-expression
Setting:	Small and large groups, formal and informal language, one-way communication, classroom
Audience:	Peers, teacher, others
Content:	Naturalistic, narrative, expressive, persuasive, expository
Responses:	Performance, skills in concert, prepared
Level:	Communication and linguistic competence

Hunter-Grundin Literacy Profiles, 1980

E. Hunter-Grundin and H. Grundin, The Test Agency, Cournswood House, North Dean, High Wycombe, Bucks, HP14 4NW, Great Britain.

The Hunter-Grundin was designed to monitor individual student progress and promote diagnostic teaching for students in grades 1-6. There are five levels and one form for each level. (We only have the information for Level 3.)

Subtests include reading, attitude toward reading, spelling, free writing and speaking. We only review the speaking subtest here.

The speaking subtest requires students to describe what is happening in a picture. Although the test is untimed, it usually takes about five minutes to complete. The teacher rates the performance in terms of confidence, enunciation, vocabulary (number of different words), accuracy of describing the picture, and imagination (going beyond what is given in the picture).

The speaking subtest has only one speaking sample in one discourse mode; there are no norms, technical information or sample, scored student responses. (The other subtests contain technical information and standards of comparisons such as norms.) There is some help with interpretation and use in the form of references to assistance in instruction. However, these references are not tied directly to performance on the speaking subtest.

Two reviews in Buros Mental Measurement Yearbook (Mitchell, 1985, 9:491) also find the instrument lacking in terms of technical information.

Purposes:	Transmitting information, narrative speaking
Setting:	One-to-one, formal language, one-way communication, classroom
Audience:	Teacher
Content:	Artificial, descriptive
Response:	Performance, skills in concert, impromptu
Level:	Linguistic and communication competence

Jones-Mohr Listening Test, 1976

J.E. Jones and L. Mohr, University Associates, 8517 Production Ave., San Diego, California 92121.

The Johns-Mohr Listening Test assesses how well people can understand spoken statements, not only by what is said, but also by how it is said. Students listen to short statements and then choose which of four meanings is implied. The test was designed for informal use by adults participating in human relations training. It could also be used with younger persons. There is one level and two forms. There are 30 items. The test takes about 25 minutes to give.

Items were pilot-tested but there is no other technical information. There are no norms, although there is some assistance with developing local norms and with assisting test takers in self-diagnosis based on results. There is a table for converting numerical scores to short descriptions (poor to excellent), but there is no rationale for these assignments. The manual provides some references to training materials, but these are not tied directly to test results.

Purposes:	Social interaction, transmitting information
Setting:	One-to-one, informal, one-way communication
Audience:	Teacher
Content:	Artificial
Response:	Multiple-choice, skills in isolation, impromptu
Level:	Communication competence

Language Proficiency Test, 1981

J. Gerard and G. Weinstock, Academic Therapy Publications, 20 Commercial Blvd., Novato, California 94947.

This instrument is designed for older students and adults (grades seven through adult) whose language skills may be low (especially persons for whom English is a second language). The purpose is to measure aural comprehension skills as well as recall of facts. The instrument is designed to determine the level of language proficiency of an ESL student and may be helpful in placement decisions. (We only review the oral/aural subtests here.)

There is one form. It includes three subtests for oral/aural skills -- commands (individually administered, requiring a physical response to one step directions); short answers (individually administered, requiring a short oral response to various questions such as "What's your favorite subject?" and "What time did you get up this morning?"); and comprehension (individually administered, requiring students to listen to a short expository passage and answer five questions requiring recall of information and aural comprehension of the question).

The emphasis is on linguistic competence because the questions were designed to reflect increasing difficulty in grammar and vocabulary. Most responses require some coordination of

skills (listening, understanding and speaking) in order to respond correctly. However, the instrument is scored only on whether the response was correct or not.

These subtests contain 25 questions and are teacher-paced. There is no information about reliability and validity, and no norms. There is a procedure for converting scores to need for placement, but there is no rationale provided for these conversions. Case studies are provided in order to assist interpretation and use. Teaching aids are referenced, but these are not tied to results.

One review in Buros Mental Measurement Yearbook (Mitchell, 1985, 9:588) states "The use of this instrument to make decisions about students in any form cannot be recommended until adequate validity and reliability evidence is provided. At that point, because of the small behavior sample for most of the subtests, only very general screening functions can be recommended."

Purposes:	Transmitting information
Setting:	One-to-one, one-way communication, formal
Audience:	Test administrator
Content:	Functional, expository, artificial
Response:	Oral, short answer; skills in concert
Level:	Linguistic competence

Listening Comprehension, Grades 1-3, 1976

S. Hohland and B. Cheney-Edwards, Educator's Publishing Service, Inc., 75 Moulton St., Cambridge, Massachusetts 02138.

This package includes several informal inventories for classroom teachers to use to assess listening in grades 1-3. The seven inventories include teacher checklists, multiple-choice tests, and a free response measure of following directions (simple performance tasks), sequencing (marking an answer sheet on the order of things in the story), using context in listening (cloze format), finding main ideas (best title for a story), forming sensory images from oral descriptions (listening to a poem and painting a picture), identifying mood and emotions, and making inferences. The inventories were designed to minimize the need for responses that do not require other than listening skills. There is an accompanying booklet of games and activities that can be used to strengthen skills in the areas assessed.

The seven skills sheets have about 75 items. There are no estimates of the amount of time needed for each activity. No technical information is provided.

Purposes:	Transmitting information, appreciation/entertainment
Setting:	One-to-one, formal language, one-way communication, classroom
Audience:	Teacher
Content:	Artificial
Responses:	Multiple-choice, drawing, physical response; mostly skills in isolation, impromptu
Level:	Linguistic and communication competence

Listening: It's Impact At All Levels on Reading and the Other Language Arts, 1979.

S.W. Lundsteen, National Council of Teachers of English, 1111 Kenyon Road, Urbana, Illinois 61801. Also ERIC ED 169 537.

This document includes several informal checklists and rating forms for classroom use. These include the Checklist of Listening Roadblocks (a self-analysis of listening problems), Coding Sheet for Teacher Behavior and Coding Sheet for Student Behavior (to be used together to analyze tapes of classroom discussions). No grade levels are indicated, but they appear to be adaptable to any grade level. No technical information is provided. No sample classroom discussions are provided to illustrate scoring.

Purposes:	Transmitting information
Setting:	Small group, classroom, informal and formal, interactive communication
Audience:	Peers, teacher
Content:	Naturalistic
Responses:	Performance, self-rate, skills in concert and isolated
Level:	Communication competence

Listening Skills Schoolwide, 1982.

T.G. Devine, National Council of Teachers of English, 1111 Kenyon Road, Urbana, Illinois 61801. Also ERIC ED 219 789.

This document provides lots of instructional and some assessment ideas for classroom teachers in the area of listening. Listening Behaviors and Habits is a teacher checklist that can be used over time to see how student behavior is changing in ten key areas (e.g. "Is less attention paid to fellow students than teacher?" "Does he/she take notes?") There are two checklists for appraising other specific listening skills and behaviors over time. There is a final checklist for appraising critical listening growth of students. The checklists are ungraded, but appear to be useful in grades 5 and above. No technical information is provided.

Purposes:	Transmitting information, analyzing messages, social interactions
Setting:	Small group, classroom, one-way and interactive communication
Audience:	Peers, teacher
Content:	Naturalistic
Responses:	Performance, skills in concert
Level:	Communication competence

Repairs Of Misunderstandings During Communication, 1979

L.C. Lee and S. Speiker, ETS Tests In Microfiche #009902, ETS Test collection, Educational Testing Service, Princeton, New Jersey 08540.

This instrument was designed to describe certain kinds of communication problems that can occur between children in free play interactions and the ways in which the children try to resolve these communication problems. There is one level and one form. The instrument was designed for PreK. Play sessions are video-taped, coding is done from a transcript of the play session. The coding procedure is quite complex and entails the following features: noting when an unclear statement occurs and coding the children's efforts at clarification including how long it takes, digressions, enunciation, the clarification strategy used, non-verbal actions, and appropriateness. The scoring rubric is described in detail, but it is not illustrated with any student transcripts.

No technical information is provided. This instrument would take a great deal of training to use properly. More information would have to be obtained from the author in order to use the instrument.

Purposes:	Social interaction
Setting:	One-to-one, informal language, interactive communication
Audience:	Peers
Content:	Naturalistic, functional
Response:	Performance, skills in concert, impromptu
Level:	Communication competence

Speaking Skills: Report 3. Assessing Student Progress on the Common Curriculum Goals, 1988

V. Spandel, Oregon State Department of Education,700 Pringle Parkway S.E., Salem, Oregon. Also ERIC ED 298 518.

This paper was written to provide assistance to school districts in Oregon on the assessment of the speaking skills in the state essential competencies. Several sample instruments are provided that are taken from other sources. Most of these are included elsewhere in this Guide. Those that aren't, include a teacher checklist and a peer-evaluation form for assessing a speech, two open-ended peer-evaluations of a discussion, and a self-evaluation of conversation skills. All of these instruments are intended for informal, classroom use.

No grade levels are provided, but it appears that the instruments would be useful in grades five and above. No technical information is provided. No sample discussion or speech topics are provided, and no sample speech or discussion transcripts are provided to illustrate scoring.

Purposes:	Transmitting information, social interactions
Setting:	Small group, one-on-one, formal and informal, interactive and one-way communication, classroom
Audience:	Peers, teacher
Content:	Naturalistic
Responses:	Performance, self-rating, peer-rating, skills in concert, impromptu and prepared
Level:	Communication competence

Test Of Implied Meanings (n.d.)

Ed Ragozzino, 671 Startouch Drive, Eugene, Oregon 97405.

The Test of Implied Meanings was designed for use at the college level. It could be used effectively at lower grade levels. The instrument attempts to measure how well the test taker understands what is said, using the *way* the words are said (the implied meaning) as well as the literal meaning of the words. A cassette tape is played and students mark which of four meanings they feel is implied by the way a statement is read. There is one form of 40 items, all multiple-choice. The test was designed for informal, classroom use. A cassette tape recorder is needed. There is no technical information. The test is easy to administer and takes about 15 minutes.

Purposes:	Social interaction, transmitting information
Setting:	One-to-one, informal, one-way communication
Audience:	Teacher
Content:	Artificial
Responses:	Multiple-choice, isolated skills, impromptu
Level:	Communication competence

SHORT REVIEWS --
EDUCATIONAL AGENCY ACTIVITIES

Although often innovative and of high quality, the products of districts, states, provinces and other educational agencies are included as short reviews because they often are not readily available -- materials usually must be requested from the agency itself. In addition, many of the documents contain instruments designed only for informal classroom assessment, or are described in documents that are meant more as reports of results or technical reports than as manuals designed for the use of others.

Although most educational agencies are happy to share their efforts, it can become burdensome to the agency to provide copies of materials to others. We urge you to request materials from educational agencies only after careful consideration of what is really needed.

For the following summaries we describe context and rate reliability and validity only for instruments that have been more formally developed. We do not rate help with selection or help with interpretation and use.

British Columbia Ministry of Education -- Enhancing and Evaluating Oral Communication in the Primary, Intermediate and Secondary Grades, 1988.

British Columbia Ministry of Education, Victoria, British Columbia, Canada. V8V 2M4

This is a package of three handbooks designed to assist classroom teachers to plan and monitor oral language learning across the curriculum. Instructional and assessment strategies are provided for affective behaviors, language awareness, listening comprehension, speech communication, critical and evaluative behaviors, interpersonal strategies and oral language codes. The handbooks include a large number of ideas for rating forms, checklists, interviews, conferences, anecdotal records, self-reports and writing to assess listening for grades K-12. None of the instruments has been pilot-tested. Some of the instruments would require knowledge and training to use.

Calgary School District -- Listening Profile and Listening Awareness Assessment Questionnaires, 1988.

Calgary School District, Calgary, Alberta, Canada. Also Journal of the International Listening Association, 2, 33-52, 1988.

The journal article referenced above describes the Edmonton and Calgary listening projects and the assessment instruments used and developed by them. Two locally developed instruments are reproduced in the article. The Listening Awareness Assessment Questionnaire (LAAQ) requires students to tape record responses to six questions about their listening behaviors and skills. These open-ended responses are categorized to further student understanding of their listening needs. (The categorizing scheme is currently under development and was not presented in the article.) There are two sets of questions; one for use at the elementary grades and one for use in junior high.

The Listening Profile is a checklist of listening behaviors that teachers use during regular class activities. It was developed from teacher anecdotal records. Ratings are done in the areas of nonverbal responses, verbal responses and behaviors. It was developed for use at grades 2, 4 and 6.

The information presented in the article is very brief. Additional information would have to be requested for proper training and use. No technical information is provided.

Glynn County, Georgia -- Oral Communication Assessment Program (1981).

D. Rubin and R.E. Bazzle, Glynn County School System, Brunswick, Georgia 31521.

This document describes the development of a speaking assessment tool to be used for judging minimum competency for high school graduation. Two tasks were developed to reflect the types of oral communication necessary in daily life -- a job interview and a public hearing.

The job interview requires students to fill out a job application form and then verbally respond to 25 questions about their qualifications, experiences and interests. Responses are multiple-choice, short answer and extended narratives. Performance is rated analytically in terms of performing social rituals, responsiveness, informativeness, initiative, interpersonal manner, language style, oral expression, speech rate and volume, and gestures.

The public hearing requires students to testify in front of a simulated school board in favor of or opposition to one of three proposals selected by the student. Responses are rated analytically in the areas of introduction, position, reasons, organization, conclusion, language style, vocal delivery and gestures. Examples of the types of statements that would receive various ratings are given. Two raters judge each performance.

Internal consistency reliabilities for the public hearing ranged from .82 to .88; those for the interview ranged from .68 to .92, depending on the rater. Using the interview and the public hearing as alternative forms, the correlation between scores for individual students was .70. Performance on the three public hearing topics was not significantly different, demonstrating that the three topics are of equal difficulty for students. Interrater reliabilities for the various tasks and occasions ranged form .72 to .87. These reliabilities are rated "fair" to "good."

Several days prior to administration, students receive and discuss guides to each of the tasks. These guides acquaint students with the importance of the communication represented in the task and the criteria by which performance will be judged.

The authors freely discuss the limitations of the instruments -- they only sample some of the skills from the domain of communication competence. Evidence for validity include: (1) content and ratings based on accepted theory; (2) review of content by experts; (3) results correlated highly with teacher judgment of student communicative ability; and (4) scores that distinguish between groups previously identified as having different achievement levels. This is rated "good."

Rubin and Mead (1984) report that the test "represents effort at speech performance assessment. The measure attempts to create a sense of context. However, a single speech sample...is not representative of general speaking skills" (p. 55).

Copies of the instructions to students, administration procedures, questions to ask students, and rating forms are included in the document cited. Complete assistance with scoring is not included in the document. This instrument may be out of print from the school district. We obtained a copy from the first author, Donald Rubin at the University of Georgia, Athens.

Purpose:	Transmitting information, analyzing messages, social interaction
Setting:	Small group and one-to-one, formal language, one-way and interactive, classroom
Audience:	Teacher, peers
Content:	Artificial, persuasive, expository, functional
Responses:	Performance, skills in concert
Level:	Communication competence

Hawaii State Department of Education -- Competency-Based Measures For Grade 3 Performance Expectations, 1987; Grade 10 CBM Technical Report, 1988.

Selvin Chin-Chance, Hawaii State Department of Education, 3430 Leahi, Building E, Honolulu, Hawaii 96815.

Hawaii's Competency Based Measures (CBMs) are designed to measure eight Foundation Program Objectives including basic skills, self-concept, problem solving, health, government and social responsibility. These are assessed by both paper and pencil multiple-choice tests and teacher ratings. Oral communication is assessed by classroom teachers based on their knowledge of the student -- no special communication situation is set up.

Ratings are done in three areas in grade three -- using and responding to language, asking questions and participating in class discussions. There are 10 ratings in grade 10 in the areas of adapting speech to informal and formal situations, adapting language for the audience, contributing to the completion of a task through a group discussion, and giving and responding to oral directions, descriptions, nonverbal messages and common visual symbols. All areas are rated on a five point scale.

A pilot test of the grade ten instrument indicated that interrater reliability is low without training and that it took teachers less than five minutes to rate each student.

Purpose:	Transmitting information, analyzing messages, social interaction
Setting:	All size groups, formal and informal language, one-way and interactive communication, classroom and playground
Audience:	Teachers, peers
Content:	Naturalistic
Response:	Performance, skills in concert, impromptu and rehearsed
Level:	Communication competence

Illinois State Department of Education -- Speaking and Listening Activities in Illinois Schools, 1986; Write On Illinois! Volume II, 1987.

Illinois State Board of Education, 100 N. First St., Springfield, Illinois 62777.

Speaking and Listening Activities describes the speaking and listening objectives that should be attained by Illinois students at the end of grades 3, 6, 8 and 11. For speaking, these include clear and expressive speaking, orderly presentation of ideas, development of ideas, use of appropriate language, nonverbal skills and use of language for a variety of purposes. Listening objectives include factual recall, identifying sequence of ideas, making inferences, identifying purposes and points of view and responding appropriately.

Informal classroom assessment ideas are also provided. For speaking these include checklists, rankings and ratings for classroom conversations, extended monologues, a job interview, and dramatic interpretation. For listening, sample passages and questions (both multiple-choice and open-ended) are provided. These are presented more as illustrations of possibilities than as actual recommendations. Issues in assessing speaking and listening are also discussed. None of the sample instruments have been pilot tested.

Write On Illinois! mainly discusses the state writing assessment. There is a brief section that addresses how the writing assessment procedures can be adapted to speaking and listening. Users need to be very familiar with the writing assessment procedures in order to adapt them to speaking or listening.

Iowa State Department of Education -- A Guide to Developing Communication Across the Curriculum, 1989; A Guide to Curriculum Development in Language Arts, 1986.

Iowa Department of Education, Grimes State Office Building, Des Moines, Iowa 50319.

These handbooks, designed for classroom teachers, take a whole language and communication competence approach. Topics discussed include: the functions of communication, strategies for designing integrated language arts learning experiences based on the various functions, and a sample procedure teachers can use to monitor students' communication abilities. This sample procedure is basically a structured log based on naturalistic observation in the classroom. It combines informal observational assessment with lesson planning. The procedure is designed for all grade levels. No technical information is provided. The documents also include an extensive bibliography and a short section on characteristics of good assessment.

Massachusetts State Department of Education -- State Speaking Assessment Instrument, Reliability and Bias Study, 1982; Development of the State Speaking Assessment Instrument, Reliability and Feasibility Study, 1983; Massachusetts Test of Basic Skills: Listening (n.d.)

Massachusetts Department of Education, Quincy Center Plaza, 1385 Hancock Street, Quincy, Massachusetts 02169.

The *Massachusetts State Speaking Assessment Instrument* was designed for students in grade 8. There are four speaking tasks: describe something liked (description), get help in an emergency (emergency), present procedures or steps in how to do something (sequence), and convince someone of a point of view (persuasion). Performance is rated in the areas of delivery (volume, rate and articulation), language (grammar and vocabulary), content, and organization. Each area is rated on a scale of 1-5 resulting in a total score of 4 to 20 for each task and 16 to 80 for the total test. The reports cited provide four parallel sets of prompts and a general overview on scoring, training raters and administering the test; however, the report is not explicit enough to reproduce either their training methods nor the assessment. Additional information would need to be requested. The authors report that an efficient rater can rate about four students per hour. Other information on costs and side benefits is presented.

Interrater reliability based on the testing of 1,014 students in 1982 was .75. In 1983, with a change in rater training, raters were within 16 points of each other 98% of the time and the consistency of pass-fail decisions was 80%. This is "fair." The developers feel that this reliability was too low to have ratings based on just one rater. Therefore, they recommend that each performance be scored by two raters.

An additional study in 1982 examined the effect of rater and student ethnicity on average ratings. In most cases the average scores for various pairs of raters did not deviate drastically from the overall average for students in various ethnic groups. In 1983, developers looked at the effects of testing occasion, rater ethnicity and rater "drift." None of these factors made a large difference in ratings. Evidence of validity is rated "fair" - "good."

Purpose:	Transmitting information, analyzing messages
Setting:	Small group, formal language, one-way communication, classroom
Audience:	Teacher, peers
Content:	Artificial, descriptive, persuasive, functional, expository
Response:	Performance, skills in concert, impromptu
Level:	Communication competence

The *Listening Assessment* addresses the eleven state listening objectives. Some of these objectives deal with general listening skills that apply to all listening situations, while others deal with specific listening situations, e.g., survival words used in emergency situations. No grade level is specified in the materials we received. There are two forms and one level.

Six passages of various types (descriptions of events and experiences, emergency messages, persuasive messages and sequences of directions) are played on tapes. Students answer 22 multiple-choice questions (both played on the tape and written in test booklets) that cover recognizing words and phrases, identifying problems, understanding words and ideas, identifying main ideas, associating details, understanding purpose, and drawing conclusions.

Since there are six passages and 22 questions there is some memory load. Internal consistency reliability is reported as .76. This is "fair." No information about validity was reported in the materials we received.

Rubin and Mead (1984) report that items were reviewed by a panel of judges, item statistics were generated and the authors determined the test is not ethnically biased. They conclude that "the test samples a variety of important listening situations and skills....The only significant drawback is the failure to test listening in a interactive context.

Purpose:	Unclear. From the students' perspective, the purpose is probably exchange of information. From the developer's perspective, the purpose might be the implied purposes in the individual passages.
Setting:	Unclear. From the students' perspective, the purpose might be one-to-one, formal language, one-way communication. From the developer's perspective, the purpose might be that implied by the passage.
Audience:	Unclear. From the students' perspective, it might be the teacher. From the developer's perspective, it might be an audience implied by the passage and question.
Content:	Artificial, descriptive, expository, persuasive, functional
Responses:	Multiple-choice, skills in isolation, impromptu
Level:	Communication competence

Michigan Department of Education -- Technical Report for the Objective Referenced Test for Critical Listening, 1980.

Michigan State Department of Education, P.O. Box 420, Lansing, Michigan 48902.

The test is designed to assess critical listening at grades 4, 7 and 10. Critical listening includes the following objectives: factual recall and identifying main idea, best summary, purpose, cause/effect, inferences, fact v. opinion, and plot. Passages include stories, informational selections, interviews, descriptions and personal narratives. Each level contains 24 items. Since there are from one to three questions on each passage, memory load is moderate.

Development included review of passages and items by educators in Michigan and pilot testing. Item statistics and complete texts of the tests are provided in the report. No other technical information is provided.

Purpose:	Unclear. From the students' perspective, the purpose is probably exchange of information. From the developer's perspective, the purpose might be the implied purposes in the individual passages.
Setting:	Unclear. From the students' perspective, the purpose might be one-to-one, formal language, one-way communication. From the developer's perspective, the purpose might be that implied by the passage.

Audience:	Unclear. From the students' perspective, it might be the teacher. From the developer's perspective, it might be an audience implied by the passage and question.
Content:	Artificial, narrative, expository, functional
Response:	Multiple-choice, skills in isolation, impromptu
Level:	Communication competence

New Hampshire State Department of Education -- Listening Skills Assessment: Manual and Script, 1980.

New Hampshire Department of Education, Division of Instruction, 64 N. Main St., Concord, New Hampshire 03301. Also ERIC ED 236 651.

This test was designed to assess listening ability in grades 5-12. There are two forms. The short form has 30 items and is designed for grade 5-8. The long form has the same 30 items plus 15 more and is designed for grades 9-12. The full form takes about 45 minutes to give.

The test requires students to listen to passages from real life (e.g., conversations, radio reports, directions, and a semi-formal talk) and answer questions that cover recall of facts, following directions, recognizing a speaker's purpose, critical listening and inferences. Some questions require both a multiple-choice and a short answer response. All passages and questions are read to the students. There are up to nine questions for each passage. There is, thus, a large memory load. There are a few instructional ideas provided, but they are not tied directly to results.

No technical information is provided.

Purpose:	Unclear. From the students' perspective, the purpose is probably exchange of information. From the developer's perspective, the purpose might be the implied purposes in the individual passages.
Setting:	Unclear. From the students' perspective, the purpose might be one-to-one, formal language, one-way communication. From the developer's perspective, the purpose might be that implied by the passage.
Audience:	Unclear. From the students' perspective, it might be the teacher. From the developer's perspective, it might be an audience implied by the passage and question.
Content:	Artificial, expository, functional
Response:	Multiple-choice, short answer, skills in isolation, impromptu
Level:	Communication competence

New York State Department of Education -- New York State English Language Arts Syllabus K-12, 1988; New York State Regents Comprehensive Examination in English, 1989.

New York State Department of Education, The University of the State of New York, Albany, NY 12234.

The *New York State English Language Arts Syllabus K-12* outlines general criteria for an effective integrated curriculum in English language arts. Accompanying the syllabus are three support manuals: *Listening and Speaking in the English Language Arts Curriculum K-12, Composition in the English Language Arts Curriculum K-12*, and *Reading and Literature in the English Language Arts Curriculum K-12.*

The syllabus document suggests the instructional objectives that need to be addressed and provides direction for the evaluation of student progress and program effectiveness. Each section examines different aspects of communication; each directs attention to the purposes, objectives, and focus skills of that particular aspect of communication, all of which support the development of interactive, interdependent, and mutually reinforcing processes that are necessary to understand the express meaning.

The *Listening and Speaking* manual contains sections on: listening and speaking in the English Language Arts curriculum, the roles of effective listeners and speakers in the communication process, the classroom as a communication environment, integrating listening and speaking across the curriculum, expectations for students K-12, and evaluation of listening and speaking skills. Several informal rating forms, checklists, peer-evaluation forms and self-evaluation forms are provided.

The *Regents Examination in English* is a comprehensive examination designed for average and above-average students. It has sections on listening, spelling, vocabulary and reading comprehension. Also, two pieces of writing are required -- a literature essay and a composition on a given topic.

The listening section consists of a three to four minute-long passage read to students by a teacher. Students listen to the first reading. They then read ten multiple-choice test items based on the passage and mark their initial selection of the correct answers. During a second reading, students can mark their answers. Test items require students to listen for essential information or facts, discover patterns, understand special use of language, formulate judgments about content, draw conclusions and understand inferences.

We received no information about the conceptual basis for the questions or the technical aspects of the test. Reading ability may be confounded with listening skill. Rubin and Mead (1984) report that the test has one passage and about 10 questions. They feel that this is too small a sample of listening performance to be treated as a separate measure, but that it is useful as part of the overall measurement of English language arts ability.

New Zealand Council For Educational Research -- Progressive Achievement Tests (PAT): Listening, 1971.

New Zealand Council For Educational Research, P.O. Box 3237, Wellington, New Zealand (04) 847-939.

The PAT-Listening test is part of an achievement test battery that also assesses reading and mathematics. It appears to be designed for students aged 8 through 18. The purpose of the listening test is to assist teachers in determining the levels of development attained by their pupils for purposes of instructional planning. It is intended that the test be given at the beginning of the school year. There are two forms.

The tests require students to comprehend and draw inferences about extended passages of orally presented material. The passages are designed to reflect situations commonly encountered by children in and out of the classroom. These include poems, directions for doing things, stories, informational pieces, descriptions of events, conversations, discussions, and radio reports. Questions include recall of facts, sequence of events, main idea, figures of speech and inferences of various types.

For all except the first level, students answer in a multi-level booklet in which different age groups begin and end at different points. Answer choices, but not questions are printed in the multi-level booklet. There are 129 questions across all levels; no one group receives more than 100. Level 1

students have a separate, disposable answer booklet. There are 42 questions. Each passage has five to seven questions. This presents a large memory load.

Internal consistency reliabilities ranged from .78 to .91 depending on form and level (median = .82). Equivalent forms reliability ranged from .71 to .83. These reliabilities are "fair" to "good."

Validity considerations include: (1) content based on common listening situations as identified by researchers and other educators; (2) pilot testing; (3) no appreciable differences in student scores when different speakers were used; (4) increase of scores with age; and (5) moderate correlations with other ability and achievement tests. This evidence is rated "fair."

Assistance with interpretation includes norms, identifying students needing special assistance, instructional ideas, and predicting reading level from listening scores. The proper cautions about overreliance on any single source of information is provided. Help with assistance is rated "fair" mainly because of the age of the norms.

There are no reviews of this instrument in Buros, Hammill, et. al. (1989) or Keyser and Sweetland (1987). The test is due to be revised in 1993.

North Carolina Department of Education -- Communication Skills, Grades 1 and 2 Assessment, 1989.

North Carolina Department of Public Instruction, 116 W. Edenton St., Raleigh, North Carolina, 27603-1712, (919) 733-3703.

This handbook is designed for use by classroom teachers in grades one and two to informally assess student progress on the North Carolina State Communication Skills. There are three parts to the assessment procedure. The first is a set of checklists covering speaking, oral language, orientation to print, listening, silent reading comprehension and unassisted writing. These are to be completed three times a year after several weeks of general observation. The second part is a checklist that focuses on communication in actual use and includes thinking skills and attitudes toward school. An attempt has been made to link speaking and listening, reading and writing. This is also intended to be used three times a year. The third part is a checklist that reflects communication skills. This is recommended for use twice a year. No technical information is provided.

Ohio State Department of Education -- Ohio English Language Arts Curriculum, 1985; Integrating Language Arts, 1985

Ohio Department of Education, Division of Elementary and Secondary Education, 65 South Front Street, Room 1005, Columbus, Ohio 43266-0308, (614) 466-2211.

These two handbooks were designed to update teachers and administrators on recent research and sound instructional practices that promote the integration of the language arts areas. The handbooks also provide guidance in developing curriculum documents that contain goals and objectives reflecting best practices and meeting state requirements. Assistance with assessing speaking and listening includes issues, criteria for evaluating instruments and skills that should be covered. The handbooks do not provide sample instruments.

Ontario Ministry of Education -- The Ontario Assessment Instrument Pool: English II Intermediate Division, 1981.

Ontario Ministry of Education, Publications Centre, 880 Bay Street, 5th Floor, Toronto, Ontario, Canada M7A 1N8.

This document presents classroom activities to develop and informally assess a number of speaking and listening skills in grades 7-10. Included are objective tests, checklists, short answer formats, self-ratings and teacher-ratings for group discussions, oral presentations, listening comprehension and language mechanics/usage. Prompts and scoring criteria for some of the instruments are provided. No technical information is available for any of the instruments.

Oregon State Department of Education -- Integrated Assessment Model: A Project-based Approach, 1988; Procedures for Assessing Listening Skills, 1984; Speaking Skills: Assessing Student Progress on the Common Curriculum Goals, 1988; Listening Skills: Assessing Student Progress on the Common Curriculum Goals, 1988; Assessing Speaking Skills: Training for Raters, 1989.

Oregon State Department of Education, 700 Pringle Parkway S.E., Salem, Oregon 87310-0290, (503) 378-8471.

The *Integrated Assessment Model* outlines one possible approach to assessing some of the more difficult to measure objectives in the Oregon Common Curriculum Goals for grades 3, 5, 8 and 11. The authors propose that students prepare a research project in which they plan, gather information, and deliver oral and written presentations of results. This allows skills to be observed during real-life tasks which require skills to be used in concert to produce a final product. Rating systems and checklists are provided for each stage -- planning, preparation and delivery. Sample research projects for the students to undertake are proposed. The procedure has not been pilot tested.

Purpose:	Transmitting information
Setting:	Small group and one-to-one, formal language, one-way communication, classroom
Audience:	Teacher, peers
Content:	Naturalistic
Responses:	Performance, skills in concert, rehearsed
Level:	Communication competence

Speaking Skills and *Listening Skills* are companion pieces designed to assist districts in complying with the state requirements of using student status on the state's Common Curriculum Goals to assist in making decisions about instruction. Each handbook contains a listing of relevant speaking or listening goals, discusses what would constitute acceptable assessment practice, provides sources for assessment help, and supplies samples of informal, classroom assessment tools from a variety of sources. For speaking, these tools include teacher rating forms, teacher checklists, peer review instruments for looking at extended monologues, and self and peer evaluations for group discussions. The listening handbook includes a multiple-choice test in response to taped information, student self-rating checklists, listening guides for students to use when listening to others and teacher checklists. None of the instruments has been pilot-tested by Oregon. Recommended ages for the instruments are not given, but they appear to be appropriate for grades 5 and above.

Procedures for Assessing Listening Skills is a package of 24 informal assessment tools designed for classroom teachers in Oregon to assess student progress toward meeting the Oregon essential competencies in grades K-8. Activities include answering questions (multiple-choice and open-

ended) about passages read, following directions (paper/pencil and performance), discriminating between sounds, auditory memory, critical listening, and non-verbal communication. Scoring criteria are included. However, not all materials necessary to give the tests are included in this document. No technical information is provided for any of the measures. Most of these measure skills in isolation.

Assessing Speaking Skills was developed by an Oregon school district (Salem-Keizer) and is distributed by the state department of education. The document is designed as a training manual for raters. The testing procedure involves giving students (grades 9-12) a choice of narrative or expository topics on which to speak. Guidelines are provided for local development of prompts. (Five sample prompts are provided.) Speeches are rated analytically on organization, delivery and language using a five-point scale. The same scale is used regardless of topic. Students are given a week to prepare their speeches. Detailed criteria for ratings, sample student speeches and instructions for students are provided. Training tapes would have to be requested separately. No technical information is provided.

Purpose:	Transmitting information
Setting:	Small group, formal language, one-way communication, classroom
Audience:	Teacher, peers
Content:	Artificial
Response:	Performance, skills in concert, rehearsed
Level:	Communication competence

Pennsylvania Department of Education -- Speech in the Classroom: Assessment Instruments, 1980.

S. Koziol, K. Cercone and E.W. Miller, Pennsylvania Department of Education, P.O. Box 911, Harrisburg, Pennsylvania 17126.

This is a package of instruments developed by the state for use by classroom teachers. The first instrument is a procedure to holistically rate students on a story that incorporates a picture prompt (supplied by the teacher). This would require training to use. A second instrument is a survey having two levels (grades 1-6, 15 questions; grades 4-12, 25 questions) which ask students and teachers to indicate which speaking activities take place in the classroom. A third instrument assesses student attitudes about various speaking activities. There are two levels -- grades 1-6, 12 questions; grades 4-12, 20 questions. No technical information is available.

Saskatchewan Provincial Department of Education -- Saskatchewan English Language Arts Curriculum, 1989.

Saskatchewan Education, 2220 College Avenue, Regina, Canada, S4P 3V7.

Saskatchewan is currently developing curriculum guides for language arts. The document we obtained was an excerpt from their grade 3 guide. The guides will include both instructional and assessment ideas for informal classroom use. Assessment as a continuous classroom process is emphasized. Listening and speaking assessment tools will include checklists, teacher ratings and self-ratings.

SHORT REVIEWS --
MEASURES THAT EMPHASIZE LINGUISTIC COMPETENCE

The instruments in this section focus primarily on linguistic competence, defined as the ability to form correct language (grammar, syntax, vocabulary, etc.). The instruments included do not represent all those available. We have selected a few representative measures for purposes of comparison to those that focus more on communication competence. **We do not rate the validity of these instruments because they represent a different construct than that presented in the rest of the Guide.**

The Fullerton Language Test for Adolescents (1986)

Arden R. Thorum, Consulting Psychologists Press, Inc., 577 College Avenue, Palo Alto, California 94306.

The Fullerton was developed to assist educators to distinguish normal from language-impaired adolescents. The test is designed for ages 11 through 18. There are eight subtests that cover blending sounds and syllables to form words, knowledge of the meaning of prefixes and suffixes, following directions having various levels of syntactic complexity, distinguishing the meaning of words that sound the same, listing as many of the objects of a given class as possible in 20 seconds, identifying the number of syllables in a word or phrase, identifying whether a sentence is grammatically correct, and using idioms correctly. Each of the eight subtests "assesses a specific function important to the acquisition and effective use of language skills by adolescents." The authors point out that the Fullerton does not include all important language process and production skills, just the major ones. No specific theoretical underpinnings for the test are mentioned.

There are 142 items; the test takes about 45 minutes to give. For all subtests except Oral Commands, responses are short answer and are given verbally by the student. The Oral Commands subtest requires a physical response. There are no multiple-choice. There is one form and one level. The test is not difficult to give, but some familiarity with the scoring rubrics are required. The test can be scored in two ways: right/wrong or descriptive (immediacy of response, self-correction, correct after repeat of stimulus or error).

The instrument is rated as "good" in terms of the information provided to the user to aid in proper selection and use.

Internal consistency reliabilities range from .70 to .85 for the various subtests. This is "fair" to "good." Test-retest reliabilities range from .84 to .96. These are "good" to "excellent."

The content of the test was based on a review of the literature, consultation on experts in the field and discussions with classroom teachers. Studies included the relationship between scores on the various subtests (they are moderately interrelated, indicating that they all measure the same type of thing); and the difference between scores of a "normal" and a "special education" population (all results were significantly different). The instrument has been used in a number of research studies.

There is considerable assistance with interpreting and using results including sample student performances, discussion of what each subtest means, average performance for various ages, and suggestions for remediation. The norms are old (1978-1979) and are based on a relatively small population (762 students in seven age ranges). Because of the norms, help with assistance and use is rated as "good" instead of "excellent."

There are no reviews of the 1986 edition in Hammill, et al. (1989) or Keyser and Sweetland (1987). Rubin and Mead (1984) reviewed the previous version and concluded "the test measures only a limited type of listening ability" (p.49). This corresponds to our placement of the instrument into the linguistic competence section. One review in Buros Mental Measurement Yearbook (Conoley and Kramer, 1989, 10:123) reports "the Fullerton appears to be a carefully developed test of adolescent language performance that is easy to administer and capable of identifying students with language impairments that may be related to academic difficulties. Suggestions for interpreting test performance into plans for language therapy make the Fullerton a particularly useful tool."

The Language Inventory For Teachers (LIT), 1982

A Cooper, and B.A. School, Academic Therapy Publications. 20 Commercial Blvd, Novato, California 94947.

The LIT is a criterion-referenced test which was developed to assist teachers to develop IEPs for special education students in grades PreK-8. There is one level and one form.

The test is intended to measure several language components identified by the author as being essential: naming and identifying objects, identifying and using object properties, identifying and using events in time and space, writing legibly, identifying and using correct grammar, writing specific sentence patterns, using various language constructs, using vocabulary, discriminating between formal and informal language, and comprehending and responding in written and spoken form. The theoretical basis for this list is not provided.

The test items require identifying important and unimportant details, fact v. opinion, figurative language, main idea, sequence of events, answering various factual and inferential questions about a passage, and narrative, descriptive and expository writing. Students respond to teacher questions by pointing to pictures, providing short oral answers, editing, and writing letters, words and paragraphs. Responses are scored primarily on the extent and quality of vocabulary, syntax and grammar. This makes the LIT primarily a measure of linguistic competence. However, some exercises touch the area of communication competence. For example "Make a report on something you have read or done in the last few weeks..." which is rated on sentence structure, paragraph structure, sequence of information, level of vocabulary, introduction, summary, and content.

There are spaces to record performance on over 500 specific language skills. If the entire inventory is given, testing time is about one hour.

The instrument is rated as "poor" in terms of providing the information necessary for selection of the instrument. Missing are a description of the theoretical basis of the instrument, checklist development, reliability and validity.

No technical information is provided.

The instrument is rated "fair" in terms of help with interpretation and use. Sample IEPs are provided.

There were no reviews of the LIT in Hammill, et al. (1989) or Keyser and Sweetland (1987). One review in Buros Mental Measurement Yearbook (Mitchell, 1985, 9:587) indicates that additional information needs to be presented on the underlying theory, inventory development, validity and reliability. This reviewer also found numerous errors in the manual and the form which make certain items hard to give and score.

The Test of Adolescent Language - 2 (TOAL-2), 1987

D.D. Hammill, V.L. Brown, S.C. Larsen, and J.L. Wiederholt, PRO-ED, 8700 Shoal Creek Blvd., Austin, Texas 78758, (512) 451-3246, FAX #512-451-8542.

The TOAL-2 was developed to identify students who might benefit from intervention, determine students' strengths and weaknesses in language abilities, document student progress in language development, and use in research. It was designed for ages 12 through 18. There is one form and one level. The test is designed to assess both receptive and expressive spoken and written language. Within each area both semantics (the meaning of words and sentences) and grammar are assessed.

There are eight subtests: Listening Vocabulary requires students to pick the picture of the word that is said (multiple-choice). Listening Grammar requires the student to pick which two of three sentences have the same meaning (multiple-choice). Speaking Vocabulary and Writing Vocabulary require the student to use a word in a sentence (performance). Speaking Grammar has students repeat sentences of various levels of complexity (performance). Reading vocabulary requires students to choose a word that goes with three other words (multiple-choice). Reading Grammar has students choose which of three written sentences mean the same thing (multiple-choice). Writing Grammar has students combine short sentences into longer ones (performance).

There are a total of 240 items. Six of the subtests can be administered either in a group or individually. Speaking Grammar and Speaking Vocabulary must be administered individually. When administering the test individually, basals (five in a row incorrect) and ceilings (five in a row correct) can be used to minimize testing time. The test is untimed but usually takes from one to three hours. Items are scored right or wrong. There is a scoring guide to identify responses that are correct.

The manual is complete in presenting information necessary to select an instrument, except for cautions about what aspects of language are not covered by the test. The rating is "good" to "excellent."

Internal consistency reliabilities range from .82 to .96 for subtests and .90 to .97 for composite scores, depending on the subtest and age of students. Test-retest reliabilities ranged from .74 to .90 for subtests and .82 to .98 for composites. Interrater reliabilities ranged from .70 to .99 from a number of studies. These reliabilities are "good" to "excellent."

There is a thorough discussion of the theoretical concerns upon which the test is based. Other evidence of validity includes: (1) scores increase with age although correlations between age and score are small (this is not an unusual finding for this age group); (2) the intercorrelation of the subtests are moderate, showing that they tend to measure the same thing; (3) scores correlated moderately with those of an ability measure (this was expected by the authors since all the tasks require some level of cognitive processing); and (4) in several studies the test distinguished between normal and handicapped populations in expected ways.

A lot of assistance is given with interpreting results. The authors include reasons why students might score as they do on the various subtests, what the various scores mean, cautions in interpreting the scores, norms based on a reasonable sample size, other questions to ask during testing to increase the information obtained from the test, how to share the test results and sources to assist in developing instruction. Assistance with interpretation and use is rated as "excellent."

Hammill, et al. (1989) rate the TOAL-2 as "acceptable" in terms of reliability, validity and norms. There are no reviews in Keyser and Sweetland (1987). Two reviews in Buros Mental Measurement Yearbook (Conoley and Kramer, 1989, 10:365) different somewhat in their endorsement of the test. One reviewer feels that the validity information is not entirely convincing, while the other feels that the information is adequate. Both generally like the content and approach.

The Test of Early Language Development (TELD), 1981

W.P. Hresko, D.K. Reid, and D.D. Hammill, PRO-ED, 8700 Shoal Creek Blvd., Austin, Texas, 78758, (512) 451-3246, FAX #512-451-8542.

The TELD has one level and one form designed for children in grades PreK-1. There are 38 items given individually to students. These are intended to cover the form (phonology, syntax and morphology) and the content (encoding and decoding meaning) of language. The use of language to achieve personal goals is expressly not covered in this test. The test attempts to cover both receptive and expressive modes. The test requires the student to match a picture to a word, match a sentence to a picture, repeat words and sentences, provide short descriptions of pictures, identify synonyms, identify classes that objects are in, interpret the inferences in sentences, and make a sentence out of a list of words.

Items require responses that are multiple-choice, gestures and short answer. Items are scored 0 or 1. Students get a "1" if listed criteria for the item are met. There are 38 items. Not all students receive all items -- there are suggested places to begin testing for various aged students. Testing continues until a student misses five items in a row. There is no time limit, but testing typically takes about 15 minutes.

Information provided to the user that aids in selection is rated "good."

Internal consistency reliabilities range from .87 to .92 (median .89) depending on age. This is rated "good".

The content was based on language models that propose two dimensions -- type (form, content and use) by mode (receptive or expressive). Correlations with 6 other measures of the same constructs were moderate. The test also differentiates between students of various ages, is moderately related to measures of general ability and achievement, and distinguishes between a normal population and a population previously identified as "communication disordered."

There are norms based on a sample of between 200 and 250 students per age range. There is assistance with interpretation and use. Help with interpretation is rated "good." More assistance could be given with use in instruction.

Hammill, et al. (1989) gave the instrument an "acceptable" rating in terms of norms, reliability and validity. There is no review of the new edition in Keyser and Sweetland (1987) or Buros Mental Measurement Yearbook (Mitchell, 1985).

A new edition will be available in January, 1990.

Test Of Language Development-2 (TOLD-2) Primary and Intermediate, 1988

Donald D. Hammill, and Phyllis L. Newcomer, PRO-ED, 8700 Shoal Creek Blvd., Austin, Texas 78758, (512) 451-3246, FAX #512-451-8542.

The TOLD-2 has two levels with one form for each level. The primary level is designed for ages 4 through 8 and the intermediate level is designed for ages 8 through 12. The purposes are to identify children who are significantly below their peers in language proficiency, to determine children's specific strengths and weaknesses in language skills, to document children's progress in language skills and to use for research. The tests are designed to measure both receptive (listening) and expressive (speaking) semantics (knowledge of meanings), syntax (knowledge of grammar) and phonology (the sound of language).

There are seven subtests in the Primary level -- picture vocabulary (receptive knowledge of word meanings; multiple-choice); oral vocabulary (the child provides definitions of words); grammatic understanding (the child picks out the picture that represents a sentence); sentence imitation (repeating sentences as a measure of grammatic knowledge); grammatic completion (the examiner reads unfinished sentences and the student supplies the missing word form); word discrimination (identifying word pairs read by the examiner as being the same or different); and word articulation (ability to produce the sounds needed for English). There are 190 items. Testing is stopped when the child misses five questions in a row. The test is not timed but takes about 30 minutes to one hour to give. Items are scored as correct or incorrect (there are detailed keys for determining when a production item is correct).

There are six subtests in the elementary level -- sentence combining (measuring syntactic ability by forming one compound sentence from two or more simple sentences), vocabulary (identifying words with the same meaning, opposite meanings, or no relationship), word ordering (measuring syntactic ability through having the child reorder a series of randomly ordered words into a sentence), generals (the child explains how three words are alike), grammatic comprehension (identifying sentences that are grammatically incorrect), and malapropisms (the child identifies and corrects a word that is used incorrectly). There are 180 items. The test is untimed but takes from 30 minutes to one hour. All the subtests are given using a basal and ceiling system -- students at different ages begin at different points on the test; testing continues until students miss five in a row at the top end and get five in a row correct at the bottom end. All items are scored as right or wrong.

The manuals are rated "good" in terms of providing the information needed to select an instrument.

Internal consistency reliabilities for both levels are reported from several studies for students of different ages and language abilities. These typically were "good" to "excellent" for both subtests and total scores. Likewise, test-retest reliabilities are "good" to "excellent."

Test content was based on linguistic theory. Other evidence of validity includes: (1) ratings by more than 100 professionals which show that test content measures the theoretical constructs on which the test is based; (2) performance on the test increases with age; (3) correlations between the subtests is moderate; (4) correlations with tests of achievement are moderate to high; (5) correlations with ability measures are moderate; (6) results from a number of studies that show the tests distinguish between groups in expected ways; and (7) factor analytic studies that show that scores on the various subtests (and other measures) cluster together in expected ways.

Several forms of assistance are provided for interpreting and using results. Help is given with profiling, definition of the various types of test scores, samples of test scores for various students, what the scores from the various subtests mean, how to determine whether the difference between scores on the profile is meaningful, possible errors in measurement, how to develop local norms,

cautions in interpreting the results, and some help with instructional planning. In general, norms are based on a reasonably sized population that is reasonably representative of the nation. For some subtests, information from the 1977 norming population was combined with that from the 1987 sample. This might make the norms somewhat "easy" for these subtests. Help with interpretation and use is rated "excellent."

No review of the 1988 edition was found in Buros Mental Measurement Yearbook (Conoley and Kramer, 1989) or Keyser and Sweetland (1987). Hammill, et al. (1989) rate the instrument as "acceptable" in terms of norms, reliability and validity.

Utah Test of Language Development-3 (UTLD-3), 1989

M.J. Mecham, PRO-ED, 8700 Shoal Creek Boulevard, Austin, Texas 78758, (512) 451-3246, FAX #512-451-8542.

The UTLD-3 was developed to identify students aged 3-9 who might fall outside the "normal" range of language development. There is one form and one level. The test has two subtests -- language comprehension and language expression. The language comprehension subtest requires students to point to the picture that represents a word, sentence, or sequence; and provide short oral answers to questions requiring identification of correct grammar, categories of objects, and vocabulary. The language expression subtest requires the student to name objects or actions, repeat words or sentences, supply a correct grammatical form, combine sentences, define words, rhyme words, make up a sentence using a supplied word and demonstrate knowledge of idioms.

These subtests attempt to measure language meanings and grammar at each of three levels -- recognition/imitation, short-term recall/rote associations and understanding. There are 100 items, but not necessarily all items are given to each student. Rather, a system of basals (five in a row correct) and ceilings (five in a row incorrect) is used. All items are scored right or wrong based on a scoring guide for each item. The test takes about 15-30 minutes to give.

The instrument is rated "good" in terms of providing the information needed for selection.

Internal consistency reliabilities range from .76 to .91 (median = .84) depending on the subtest and age. This is rated "good."

Validity information includes: (1) content based on theories of language components and development; (2) moderate relationships with other tests that attempt to measure the same thing; (3) scores improve with age; (4) scores correctly identified a group that had already been identified as being low achieving; and (5) the subtests relate moderately with each other.

Norms based on a small population per age are provided for both reception (listening) and expression (speaking). Reception and expression scores are combined to form a "language quotient." This allows users to assess modalities separately as well as obtain an overall index of language competence. Help with interpretation and use is rated "good." Proper cautions are provided. There is little assistance with use of results beyond screening, but this was the only recommended use for the test.

No reviews of the 1989 edition are available from Buros Mental Measurement Yearbook (Conoley and Kramer, 1989), Hammill, et al. (1989) or Keyser and Sweetland (1987).

ADDITIONAL ASSESSMENT INSTRUMENTS

These instruments were not obtained in time for inclusion as long or short reviews. The following descriptions are based on information in test publisher catalogs, reviews by others, and descriptions in other research studies.

Evaluating Communicative Competence -- A Functional Pragmatic Procedure (n.d.)

Chariann S. Simon, United Educational Services, Inc., P.O. Box 605, East Aurora, New York, 14052, (800) 458-7900.

This is a series of 20 informal evaluation tasks that serve as probes of auditory and expressive language skills needed for classroom and social purposes. They cover language processing, skills in talking about language (metalinguistic skills), and functional uses of language for various communicative purposes. The instrument was designed for students aged 9-17.

A review in Buros Mental Measurement Yearbook (Conoley and Kramer, 1989, 10:110) states that administration takes about 60 minutes. Additional time is required to evaluate responses. There is some technical information. The review is positive in terms of the scope of the tasks presented.

Interpersonal Language Skills Assessment (ILSA)

Carolyn M. Blagden and Nancy L. McConnell, LinguiSystems, Inc., 3100 Fourth Ave., East Moline, Illinois 61244.

This instrument is a structured observation of children in grades 3-9 while they are playing a board game such as Sorry. The purpose is to assess students' use of the linguistic social skills necessary for successful interpersonal interaction.

Three to four students play the game. If the game is videotaped, all students can be rated. If the game is not taped, only one student is observed. The students' comments are categorized by type: advising/predicting, commanding, commenting, criticizing, informing, justifying, requesting, and supporting. There are norms for ages 8-14. Technical information is available.

Kentucky Comprehensive Listening Test, 1980.

R.N. Bostrum and E.S. Waldhart, University of Kentucky, Lexington.

From a study by Rubin and Roberts (1987): this instrument covers short term listening, short term listening with rehearsal, interpretation of meaning and lecture comprehension (long term memory). It emphasizes attention, comprehending and remembering. Distractions are built into the taped material.

Test of Pragmatic Skills, 1986.

Brian Shulman, United Educational Services Inc., P.O. Box 605, East Aurora, New York 14052, (800) 458-7900.

This instrument attempts to assess three through eight year old children's use of language to signify conversational intent. Ten categories of communicative intentions and functions are covered -- naming/labelling, reasoning, requesting information, requesting action,

answering/responding, informing, summoning/calling, greeting, closing conversation, and rejection/denial.

There are four guided play interactions with examiner probes designed to elicit the child's conversational intentions. The play situations involve puppets, pencil and paper, telephones and blocks.

There is a "Language Sampling Supplement to use if the child has successfully passed the conversational intent portion of the test. This supplement helps one assess how the child uses conversational intent to organize discourse."

The test is standardized and normed.

Two reviews in Buros Mental Measurement Yearbook (Conoley and Kramer, 1989, 10:371) report that the students receive scores on each task as well as an overall score. Scores for individual intentions are not provided. Although there is some technical information available, both reviewers feel that this could be expanded, especially since the behavior sample is somewhat limited. The general feeling is, however, that this test is a useful addition if used with caution, because there are not many standardized measures of intentional competence.

APPENDIX D

Resources

PRINT RESOURCES

Bock, D.G. and Bock, E.H. (1981). *Evaluating classroom speaking.* Urbana, Illinois: National Council of Teachers of English. Also ERIC ED 214 213.

This document provides a complete discussion of how to assess speaking (extended monologues) in the classroom. The authors include discussions of issues, what to be careful of, and how to construct an evaluation instrument. Several sample rating forms are included.

Conoley, J.C. and Kramer, J.J. (Eds.) (1989). *Tenth Mental Measurement Yearbook*. Lincoln, NB: University of Nebraska Press.

This is the 10th edition of the Mental Measurement Yearbook which reviews tests and assessment devices in a number of content areas.

Devine, T.G. (1982). *Listening skills schoolwide: Activities and programs*. Urbana, Illinois: National Council of Teachers of English. Also ERIC ED 219 789.

This document focuses mainly on instructional ideas in the area of listening. A few informal assessment checklists and rating forms are included.

Dickson, W.P. (1981). *Children's communication skills.* New York: Academic Press.

This book is an anthology by researchers in the areas of referential communication and sociolinguistics. Referential communication research usually proceeds by having persons participate in artificial communication situations in order to help explain the underlying cognitive abilities and correlates of performance. The sociolinguistic tradition seeks to understand communication in terms of the social and contextual setting in which it takes place. The book attempts to bring together these two fields, both dealing with communication.

Goodman, K.S., Goodman, Y.M. and Hood, W.J. (1989). *The whole language evaluation book.* Portsmouth, New Hampshire: Heinemann.

This anthology of essays by teachers and writing consultants explores a variety of issues and approaches relating to whole language evaluation at the classroom level. Included are samples of self and peer-evaluation as well as teacher-directed evaluation ratings, checklists, anecdotal records, and miscues. Broad topics include the theory and general principles of whole language evaluation, changes in evaluation through the grade levels, and evaluation of students who have writing difficulties. The major focus is on helping teachers make better use of evaluation to understand their students, and on integrating whole language evaluation and instruction.

Hammill, D.D., Brown, L. and Bryant, B.R. (1989). *A consumer's guide to tests in print.* Austin, Texas: PRO-ED.

This book rates about 300 tests on technical quality -- norms, validity and reliability. They are rated A (highly recommended), B (acceptable), or F (not recommended). Twelve measures of writing and 15 measures of speaking are included. Achievement test series are not included.

Joint Committee on Testing Practices (1988). *Code of fair testing practices in education*. Washington, D.C.: Joint Committee on Testing Practices, American Psychological Association, 1200 17th Street, NW, Washington, D.C. 20036.

The Code of Fair Testing Practices in Education addresses the obligations to test takers of those who develop and use tests. Standards are presented in four areas: developing/selecting tests, interpreting scores, striving for fairness and informing test takers.

Lundsteen, S.W. (1979). *Listening: Its impact at all levels on reading and the other language arts*. Urbana, Illinois: National Council of Teachers of English. Also ERIC ED 169 537.

This article provides a long, detailed discussion of the skills and abilities involved in listening comprehension. It includes several informal checklists and rating forms for classroom use.

Mitchell, J.V. (Ed.)(1985). *Ninth Mental Measurement Yearbook*. Lincoln, NB: University of Nebraska Press.

This is the ninth edition of the Mental Measurement Yearbook which reviews numerous tests of all types.

McCroskey, J.C. and Daly, J.A. (Eds.) (1986). *Personality and interpersonal communication*. Newbury Park, CA: Sage.

This book contains a number of papers on how communication anxiety and other personality characteristics affect communication.

Reed, L. (1983). *Assessing children's speaking, listening, and writing skills. The talking and writing series, K-12*. Washington, D.C.: Dingle Associates. Also ERIC 233 380.

This paper takes a classroom teacher's perspective on assessing communication skills. It describes issues, considerations and procedures for assessing writing, speaking and listening. Several informal assessment tools are included.

Tardy, C.H. (Ed.) (1988). *A handbook for the study of human communication: Methods and instruments for observing, measuring, and assessing communication processes*. Norwood, NJ: Ablex.

This book contains a number of articles on the assessment of communication processes. However, the various authors approach the topic more from a counseling/personality perspective than a cognitive skill perspective. For example, the characteries assessed include other-orientation, self-centered behavior, social adaptability, empathy, marital relationships, social composure, wit, appropriate disclosure, etc. There is some overlap with the affective instruments included in our reviews, such as communication apprehension, and occasionally the scales contain some aspects of skill in grammar and other cognitive knowledge.

Ting-Toomey, S. and Korzenny, F. (Eds.) (1989). *Language, communication, and culture: Current directions*. Newbury Park, CA: Sage.

This book presents a number of papers on the general relationship between culture, language and communication.

PROFESSIONAL ORGANIZATIONS

American Speech-Language-Hearing Association, 10801 Rockville Pike, Rockville, Maryland 20852, (301) 897-5700.

Conference on College Composition and Communication, 1111 Kenyon Road, Urbana, Illinois 61801.

International Communication Association, 8140 Burnet Road, P.O. Box 9589, Austin, Texas 78766.

International Listening Association, Dr. Charles Roberts, Executive Director. P.O. Box 90340 McNeese State University, Lake Charles, Louisiana 70609-0340, (318) 475-5120.

International Reading Association, 800 Barksdale Rd, P.O. Box 8139, Newark, Delaware 19714-8139.

National Council of Teachers of English, 1111 Kenyon Road, Urbana, Illinois 61801.

Speech Communication Association, 5105 E. Backlick Rd, #E, Annandale, Virginia 22003.

OTHER ORGANIZATIONS

National Assessment of Educational Progress, Educational Testing Service, CN6710, Princeton, New Jersey 08541-6710.

The National Assessment of Educational Progress (NAEP) conducts yearly national studies of student achievement in a variety of subject areas.

OERI Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation, (Eva Baker, Director), UCLA Graduate School of Education, 145 Moore Hall, Los Angeles, California 90024-1522, (213) 206-1530.

CRESST is involved in a number of innovative assessment projects.

APPENDIX E

Summary Table

INSTRUMENT	DESCRIPTION FOCUS	GRADES	# LEVELS	# FORMS	#ITEMS/ TASKS	AD. TIME	FORMAT
Class Apprehension About Participation Scale (1987)	Speaking Anxiety	7-13	1	1	20 items	10 min.	Questionnaire-Group
College Outcome Measures Program (1986)	Extended Monologues; Communication Competence	13+	1	3	15 tasks	2-1/2 hrs.	Performance-Individual
Communication Competency Assessment Instrument (1982)	Extended Monologue Listening Comprehension; Communication Competence	9-13+	1	1	3 tasks	30 min.	Performance-Individual
Diagnosis of Group Membership (1953)	Group Discussion; Communication Competence	9-13+	1	1	1 task	?	Performance-Group
English Language Skills Profile (1987)	General Speaking Interactive Speaking-Listening Listening Comprehension Group Discussion; Communication Competence	7-12	2	2	40 items 2 tasks	At Least 90 min.	Cloze, multiple choice, short answer, self-evaluation, performance-Group
Evaluating Communication Competence (n.d.)	?	3-12	1	1	20 tasks	60 min.	Performance-Individual
Fullerton Language Test for Adolescents (1986)	Speaking Listening; Linguistic Competence	5-12	1	1	142 items	45 min.	Short verbal answers - Individual
Hunter-Grundin Literacy Profiles (1980)	Extended Monologue; Linguistic and Communication Competence	1-6	5	1	1 task	5 min.	Performance-Individual

* Only rated for commercially available instruments.

** Research instruments are not rated in these areas since the intent of the source is to report on the use of the instrument in research not as documentation for users. These sources therefore, generally lack help with selection and use.

98

OTHER	HOW WELL* MANUAL PROVIDED INFO.	HELP WITH* INTERP.	TECHNICAL ADEQUACY RELIABILITY	TECHNICAL ADEQUACY VALIDITY	COMMENTS	AVAILABLE FROM:
	** N/A	** N/A	Good-Excellent	Good		M.R. Near Communication Ed., 36, 154-166, 1987.
Scoring takes an additional 1 hr. per student	Fair-Good	Fair	Good-Excellent	Good		College Outcomes Measures Program, ACT, P.O. Box 168, Iowa City, Iowa 52243
	*** Unknown	*** Unknown	Good	Good	Associated info. in ERIC Ed 210-748 & Commun. Ed., 31, 19-32, 1982.	Speech Communication Association, 5105 Backlick Rd., Annandale, VA 22003
	** N/A	** N/A	Unknown	Unknown		L. Crowell, Speech Teacher, 2, 26-32, 1953
	Good	Good	Good-Excellent	Fair	Requires a tape recorder and training in scoring.	MacMillan Education Houndmills, Basingstoke, Hampshire RG21 2XS, Great Britain
Extra time needed for scoring	?	?	?	?	Instrument was not obtained in time for pub. Info. comes from catalog	United Educational Services, Inc. P.O. Box 665, East Aurora, NY 14052, (800)458-7900
	Good	Good	Fair-Excellent	N/A+		Consulting Psychologists Press, 577 College Ave., Palo Alto, CA 94306
	Fair	Fair	Unknown	Unknown	Ratings only apply to the speaking subtest.	The Test Agency, Cournswood House North Dean, High Wycombe, Bucks., HP14 4NW, Grt. Britain

*** We were not able to review the manual provided with the assessment materials prior to publication deadline.

+ Instruments focusing on linguistic competence are not rated to avoid confusion with those measuring communication competence.

GENERAL INSTRUMENTS

INSTRUMENT	DESCRIPTION FOCUS	GRADES	# LEVELS	# FORMS	#ITEMS/ TASKS	AD. TIME	FORMAT
Interactional Competency Checklist (1978)	Interactive Speaking-Listening; Linguistic and Communication Competence	K-3	1	1	16 items	?	Teacher checklist-Individual
Jones-Mohr Listening Test (1976)	Listening Comprehension; Communication Competence	9-13+	1	2	30 items	25 min.	Multiple-choice-Group
Kentucky Comprehensive Listening Test (1980)	Listening Comprehension; Communication Competence	9-13+	?	?	?	?	Multiple-choice-Group
Language Communication Skills Task (1972)	Referential Communication; Linguistic and Communication Competence	K-2	1	1	2 tasks	25 min.	Performance-Dyads
Language Inventory for Teachers (1982)	Speaking and Listening; Linguistic Competence	PreK-8	1	1	500 items	?	Multiple-choice, short verbal & written answers-Individual
Language Proficiency Test (1981)	Listening Comprehension; Linguistic and Communication Competence	7-13+	1	1	25 items	?	Performance, and multiple-choice short verbal answers-Individual
Notebook Communication Game (1979)	Referential Communiction; Communication Competence	PreK-13+	1	1	12 items	?	Performance-Dyads

* Only rated for commercially available instruments.

** Research instruments are not rated in these areas since the intent of the source is to report on the use of the instrument in research not as documentation for users. These sources therefore, generally lack help with selection and use.

OTHER	HOW WELL* MANUAL PROVIDED INFO.	HELP WITH* INTERP.	TECHNICAL ADEQUACY RELIABILITY	VALIDITY	COMMENTS	AVAILABLE FROM:
	** N/A	** N/A	Unknown	Fair	Source listed does not include criteria for rating student performances or specifics about the task presented to the students.	J. Black, Research in the Teaching of Eng., 13, 49-68, 1978.
Students comprehend meaning through how a statement is read.	Fair	Fair	Unknown	Unknown	This test was expressly designed as a training device.	University Associates, 8517 Production Ave. San Diego, CA 92121
Requires a tape recorder	** N/A	** N/A	?	?	Instrument was not obtained in time for pub. Info. taken from a study in which the instrument was used.	Bostrum & Waldhard, U. of KY, Lexington KY, (606)257-7800
	** N/A	** N/A	Fair	Fair	Entire instrument is not provided in the source listed.	M.C. Wang et al., Learning Research & Development Center., U. of Pittsburgh, PA 15213
	Poor	Fair	Unknown	Unknown		Academic Therapy Pubs., 20 Commercial Blvd., Novato, CA 94947
	Fair	Fair	Unknown	Unknown	Designed primarily for non-English speakers and persons with low skill levels.	Academic Therapy Pubs., 20 Commercial Blvd., Novato, CA 94947
	** N/A	** N/A	Unknown	Unknown	Entire instrument is not provided in the source listed.	W.P. Dickson, Center For Individualized Schooling, U. of Wisconsin, Madison WI.

*** We were not able to review the manual provided with the assessment materials prior to publication deadline.

+ Instruments focusing on linguistic competence are not rated to avoid confusion with those measuring communication competence.

INSTRUMENT	DESCRIPTION FOCUS	GRADES	# LEVELS	# FORMS	#ITEMS/ TASKS	AD. TIME	FORMAT
Personal Report of Communication Apprehension (1986)	Communication Anxiety	9-13+	1	1	24 items	?	Questionnaire-Group
Profile of Nonverbal Sensitivity (1979)	Nonverbal Communication	3-13+	1	1	220 items	45 min.	Multiple-choice Group
Repairs of Misunderstanding During Communication, (1979)	Interactive Speaking-Listening; Communication Competence	PreK	1	1	1 task	?	Performance-Individual
Test of Adolescent Language-2 (1987)	Listening and Speaking: Linguistic Competence	7-12	1	1	240 items	1-3 hours	Multiple-choice short verbal answers-Individual or Group
Test of Early Language Development (1981)	Listening and Speaking: Linguistic Competence	PreK-1	1	1	38 items	15 min.	Multiple-choice Short verbal answers-Individual
Test of Implied Meanings (n.d.)	Listening Comprehension; Communication Competence	7-13+	1	1	40 items	15 min.	Multiple-choice-Group
Test of Language Development-2 (1988)	Listening and Speaking; Linguistic Competence	PreK-7	2	1	190 items	30 min.-1 hour	Multiple-choice & short verbal answers-Individual
Test of Pragmatic Skills (1986)	?	PreK-3	1	1	4 tasks	?	Performance-Individual
Two Referential Communication Tasks (1979)	Referential Comunication; Communication Competence	9-13+	1	1	Varies	Varies	Performance-dyads; Multiple-choice-group

* Only rated for commercially available instruments.

** Research instruments are not rated in these areas since the intent of the source is to report on the use of the instrument in research not as documentation for users. These sources therefore, generally lack help with selection and use.

OTHER	HOW WELL* MANUAL PROVIDED INFO.	HELP WITH* INTERP.	TECHNICAL ADEQUACY RELIABILITY	VALIDITY	COMMENTS	AVAILABLE FROM:
	** N/A	** N/A	Unknown	Unknown		J.C. McCroskey, An Introduction to Rhetorical Communication. Englewood Cliffs, NJ: Prentice Hall, 1986.
Items administered on videotape	*** Unknown	*** Unknown	Fair-Excellent	Good-Excellent		Irvington Publishers 551 Fifth Ave., New York, NY 10017 (212)777-4100
One play situation is videotaped	** N/A	** N/A	Unknown	Unknown	This instrument would take a great deal of training to use properly.	L.C. Lee & S. Speiker, ETS Tests in Microfiche #009902, ETS Test Collection, Princeton, NJ, 08541
2 of the 8 subtests must be given individually	Good-Excellent	Excellent	Good Excellent	N/A+		PRO-ED, 8700 Shoal Creek Blvd., Austin, TX, 78758 (512)451-3246
Not all students take all items.	Good	Good	Good	N/A+		PRO-ED, 8700 Shoal Creek Blvd., Austin, TX, 78758 (512)451-3246
Students comprehend meaning through how a statement is read.	** N/A	** N/A	Unknown	Unknown	A cassette tape recorder is necessary.	Ed Ragozzino, 671 Startouch Dr., Eugene, OR 97405
Not all students take all items.	Good	Excellent	Good-Excellent	N/A+		PRO-ED, 8700 Shoal Creek Blvd., Austin TX, 78758 (512)451-3246
The 4 tasks are guided play in which the rater participates.	?	?	?	?	Test was not obtained in time for pub. Info. is from publ. catalogs.	United Educational Services, Inc., PO Box 605, East Aurora, NY 14052 (800)458-7900
Number of trials varies, depending on # of students.	** N/A	** N/A	Unknown	Unknown		W.P. Dickson et al., Center for Individualized Schooling, Univ. of Wisc., Madison, WI

*** We were not able to review the manual provided with the assessment materials prior to publication deadline.

\+ Instruments focusing on linguistic competence are not rated to avoid confusion with those measuring communication competence.

GENERAL INSTRUMENTS

	DESCRIPTION						
INSTRUMENT	FOCUS	GRADES	# LEVELS	# FORMS	#ITEMS/ TASKS	AD. TIME	FORMAT
Utah Test of Language Development-3 (1989)	Listening and Speaking; Linguistic Competence	PreK-4	1	1	100 items	15-30 min.	Multiple-choice and short verbal answers-Individual
Watson-Barber High School Listening Test (1989)	Listening Comprehension; Linguistic and Communication Competence	7-12	1	2	50 items	35 min.	Multiple-choice-Group
Willingness To Communicate Scale (1987)	Willingness to Communicate	9-13+	1	1	20 items	?	Questionnaire-Group

* Only rated for commercially available instruments.

** Research instruments are not rated in these areas since the intent of the source is to report on the use of the instrument in research not as documentation for users. These sources therefore, generally lack help with selection and use.

OTHER	HOW WELL* MANUAL PROVIDED INFO.	HELP WITH* INTERP.	TECHNICAL ADEQUACY RELIABILITY	TECHNICAL ADEQUACY VALIDITY	COMMENTS	AVAILABLE FROM:
Not all students take all items	Good	Good	Good	N/A+		PRO-ED, 8700 Shoal Creek Blvd., Austin, TX, 78758 (512)451-3246
Test ad. in its entirety on audio- or video-tape.	Fair	Fair	Poor	Fair		Spectra, Inc., Box 1708 Auburn, Alabama 36831-1708
	** N/A	** N/A	Fair-Good	Fair		McCroskey & Richmond Willingness to communicate. In McCroskey & Daly (Eds.), Personality and Interpersonal Communication. Beverly Hills, CA: Sage, 1987.

*** We were not able to review the manual provided with the assessment materials prior to publication deadline.

+ Instruments focusing on linguistic competence are not rated to avoid confusion with those measuring communication competence.

ANTHOLOGIES*

	DESCRIPTION			
	FOCUS	GRADES	FORMAT	CONTENT
Assessing Children's Speaking, Listening and Writing Skills (1983)	Group Discussion; Communication Competence	5-12	Performance, Self-rating	1 rating form covering group discussions
Evaluating Classroom Speaking (1981)	Extended Monologues; Linguistic and communication competence	5-12	Performance, Teacher and Self-evaluation	10 rating forms covering various types & aspects of extended monologues
Listening Comprehension, Grades 1-3 (1976)	Listening Comprehension; Linguistic and Communication Competence	1-3	Multiple-choice, Checklists, Short response	7 inventories covering directions, sequence, main idea, sensory images, inferences, using context
Listening: It's Impact At All Levels... (1979)	Listening Problems Group Discussion; Communication Competence	1-12	Performance, Self-report Teacher Checklists	3 instruments covering listening problems & group discussions
Listening Skills Schoolwide (1982)	Listening; Communication Competence	5-12	Teacher Checklists	4 instruments covering listening skills & behaviors
Speaking Skills: Report 3... (1988)	Extended Monologues Group Discussion Interactive Speaking/Listening; Communication Competence	5-12	Teacher Checklist, Self-evaluation, Peer-evaluation	5 instruments covering group discussions, extended monologues, conversation skills

* These are articles, and other sources providing informal assessment tools for classroom use. There is typically no technical information. However, they are usually associated with many instructional ideas.

OTHER	TECHNICAL INFORMATION RELIABILITY	VALIDITY	COMMENTS	AVAILABLE FROM
Provides considerations when doing classroom assessment.	Unknown	Unknown	No sample topics for group discussions, no anchor performances to assist rating.	L. Reed, ERIC ED 233 380
Describes in detail how to do a classroom speaking assessment.	Unknown	Unknown	No sample topics for speeches; no anchor speeches to assist rating.	D.G. Bock & E.H. Bock Evaluating Classroom Speaking, Speech Communication Assoc. Annandale, VA. Also ERIC ED 214 213
Includes an accompanying booklet of games and activities to build skills.	Unknown	Unknown		Educator's Publishing Service, 75 Moulton St. Cambridge, MA 02138
	Unknown	Unknown	No sample topics for group discussions; no anchor performances to assist rating.	S.W. Lundsteen, NCTE, 1111 Kenyon Rd., Urbana, IL 61801. Also ERIC ED 169 537
Lots of instructional ideas are included.	Unknown	Unknown		T.G. Devine, NCTE, 1111 Kenyon Rd. Urbana, IL 61801. ERIC ED 219 789
	Unknown	Unknown		V. Spandel, Oregon State Dept. of Ed., 700 Pringle Parkway SE, Salem, OR. Also ERIC ED 298 518

ACHIEVEMENT TESTS

INSTRUMENT	DESCRIPTION						
	FOCUS	GRADES	# LEVELS	# FORMS	#ITEMS/ TASKS	AD. TIME	FORMAT
California Achievement Test (1985)	Listening Comprehension; Linguistic and Communication Competence	K	1	2	52 items	90 min.	Multiple-Choice-Group
CIRCUS (1976)	Listening Comprehension; Linguistic Competence	Pre-K-3	4	1	25-40 items	30-40 min.	Multiple-Choice & Short verbal Answer-Individual and Group
Comprehensive Test of Basic Skills (1989)	Listening Comprehension; Linguistic and Communication Competence	K-2	3	2	48-66 items	38-55 min.	Multiple-Choice-Group
Comprehensive Testing Program (1982)	Listening Comprehension; Linguistic and Communication Competence	1-3	2	1	40 items	40-60 min.	Multiple-Choice-Group
Diagnostic Achievement Battery (1984)	Listening Comprehension; Linguistic and Communication Competence	1-9	2	1	122 items	?	Short verbal answers; Individual
Iowa Test of Basic Skills (1990)	Listening Comprehension; Linguistic and Communication Competence	K-3	4	2	31-32 items	?	Multiple-Choice-Group
Iowa Test of Basic Skills-Listening Supplement (1990)	Listening Comprehension; Linguistic and Communication Competence	3-8	6	1	95 items	?	Multiple Choice Group
Language Diagnostics Test (1988)	Listening Comprehension; Linguistic Competence	1-9	4	1	192 items	?	Multiple-Choice-Group

* Only the listening subtests are reviewed. Although most of the tests have good norms and have been developed using standard procedures, they are generally not explicit in terms of the theoretical perspective of the listening test and they generally do not provide explicit validity information. Also, although the tests are usually very complete in terms of assistance with interpretation and use (forms, proper cautions), these are not specific to listening. Also, without an explicit theoretical base, it is difficult to interpret and use the results.

** Test was intended to measure prereading not listening, so validity is not rated.

OTHER	HOW WELL* MANUAL PROVIDED INFO.	HELP WITH* INTERP.	TECHNICAL ADEQUACY RELIABILITY	VALIDITY	COMMENTS	AVAILABLE FROM:
A supplemental listening test is the Listening Test (see below)	Fair Good	Fair- Good	Unknown	** N/A	Technical info. was not provided with the samples we received	CTB/McGraw-Hill, 2500 Garden Rd., Monterey, CA 93940, 800-538-9547
Level A has 3 extra subtests-What Words Mean, How Words Work, & Noises	Fair- Good	Fair Good	Fair- Good	Fair	Norms have not been updated since 1977.	CTB/McGraw-Hill, 2500 Garden Rd., Monterey, CA 93940, 800-538-9547
A supplemental listening test is the Listening Test (see below)	Fair- Good	Fair- Good	Fair- Good	** N/A		CTB/McGraw-Hill, 2500 Garden Rd. Monterey, CA 93940, 800-538-9547
	Fair- Good	Fair- Good	Fair	Poor- Fair	Tests are only available to members Norms are old (1973)	Educational Records Bureau, Bardwell Hall 37 Cameron St., Wellesley, MA 02181, (617)235-8920
	Fair- Good	Fair- Good	Poor- Good	Fair		PRO-ED, 8700 Shoal Creek Blvd., Austin, TX 78758
A Supplemental test is available (see below)	Fair- Good	Fair- Good	Fair	Fair		Riverside Publ. Co. 8420 Bryn Mawr Ave. Chicago, IL 60631, 800-373-9540
	Fair- Good	Fair- Good	Fair- Good	Fair		Riverside Publ. Co. 8420 Bryn Mawr Ave. Chicago, IL 60631, 800-323-9540
	Fair- Good	Fair- Good	Poor- Fair	Fair		Psychological Corp., 555 Academic Court San Antonio, TX 78204 800-228-0752

	DESCRIPTION						
INSTRUMENT	FOCUS	GRADES	# LEVELS	# FORMS	#ITEMS/ TASKS	AD. TIME	FORMAT
Listening Test (1985)	Listening Comprehension; Communication Competence	3-12	6	1	18-20 items	30-40 min.	Multiple-Choice-Group
Metropolitan Achievement Test-6 (1987)	Listening Comprehension; Linguistic Competence	K-3	4	2	10-24 items	18-25 min.	Multiple-Choice-Group
Metropolitan Readiness Test (1986)	Listening Comprehension; Linguistic and Communication Competence	K-1	2	1	18-27 items	15-30 min.	Multiple-Choice-Group
National Achievement Test (1989)	Listening Comprehension; Linguistic and Communication Competence	K-1	3	2	55-71 items	40-50 min.	Multiple-Choice-Group
National Test of Basic Skills (1985)	Listening Comprehension; Linguistic Competence	PreK-1	3	2	10-40 items	30-35 min.	Multiple-Choice-Group
Stanford Achievement Test (1989)	Listening Comprehension; Linguistic and Communication Competence	1-9	8	2	45 items	30 min.	Multiple Choice-Group
Stanford Early School Achievement Test (1988)	Listening Comprehension; Linguistic and Communication Competence	K-1	2	1	45 items	30 min.	Multiple-Choice-Group
Survey of Basic Skills (1985)	Listening Comprehension; Communication Competence	K-1	2	2	22-23 items	20 min.	Multiple-Choice-Group
Tests of Achievement and Proficiency-Listening Supplement (1987)	Listening Comprehension; Linguistic and Communication Competence	9-12	4	1	50 items	40 min.	Multiple-Choice Group

* Only the listening subtests are reviewed. Although most of the tests have good norms and have been developed using standard procedures, they are generally not explicit in terms of the theoretical perspective of the listening test and they generally do not provide explicit validity information. Also, although the tests are usually very complete in terms of assistance with interpretation and use (forms, proper cautions), these are not specific to listening. Also, without an explicit theoretical base, it is difficult to interpret and use the results.

** Test was intended to measure prereading not listening, so validity is not rated.

OTHER	MANUAL PROVIDED INFO.	HELP WITH* INTERP.	TECHNICAL ADEQUACY RELIABILITY	VALIDITY	COMMENTS	AVAILABLE FROM:
	Fair-Good	Fair-Good	Unknown	Unknown	Technical info. was not provided with the samples we received.	CTB/McGraw-Hill 2500 Garden Rd. Monterey, CA 93940 800-538-9547
3 of the 4 levels have fewer than 10 listening comp. questions	Fair-Good	Fair-Good	Unknown	Unknown	Technical info was not provided with the samples we received	Psychological Corp. 555 Academic Court San Antonio, TX 78204, 800-228-0752
	Fair-Good	Fair-Good	Fair-Good	Fair	There is a supplemental teacher checklist having 14 ratings.	Psychological Corp. 555 Academic Court San Antonio, TX 78204 800-228-0752
	Fair-Good	Fair-Good	Unknown	Unknown	Technical info was not provided with the samples we received.	American Testronics P.O. Box 2270, Iowa City, IA 52244, 800-553-0030
	Fair-Good	Fair-Good	Fair-Good	Fair		American Testronics P.O. Box 2270, Iowa City, IA 52244, 800-553-0030
	Fair-Good	Fair-Good	Unknown	Unknown	Technical info was not provided with the samples we received.	Psychological Corp. 555 Academic Court San Antonio, TX 78204, 800-228-0752
	Fair-Good	Fair-Good	Unknown	Unknown	Technical info was not provided with the samples we received.	Psychological Corp. 555 Academic Court San Antonio, TX 78204, 800-228-0752
	Fair-Good	Fair-Good	Fair	Poor-Fair		SRA, 155 N. Wacker Dr. Chicago, IL 60606.
	Fair-Good	Fair-Good	Good	Fair		Riverside Publ. Co. 8420 Bryn Mawr Ave. Chicago, IL 60631, 800-323-9540

EDUCATIONAL AGENCIES*

AGENCY	DESCRIPTION			
	FOCUS	GRADES	FORMAT	CONTENT
British Columbia Ministry of Education	Informal Classroom Assessment	K-12	Ratings, check-lists, interviews, self-evaluation, record reviews	A number of instruments covering affect, listening comprehension, extended monologues, linguistic competence and interactive communication
Calgary School District	Formal District Assessment	2,4,6	Checklist, self-evaluation	2 instruments covering listening behaviors and skills
Glynn County, Georgia	Formal Local Assessment	9-12	Multiple-choice, short verbal responses, performance.	2 performance tasks covering speaking & listening - job interview and public hearing
Hawaii State Dept. of Ed.	Formal State Assessment	3,6, 8,10	Teacher rating	1 rating form for each grade covering a variety of speaking skills.
Illinois State Dept. of Ed.	Informal Classroom Assessment	3,6,8, 11	Checklists, rankings, ratings, multiple-choice	Several instruments covering speaking & listening-job interview, classroom conversation, extended monologues, & dramatic interpretation.
Iowa State Dept. of Ed.	Informal Classroom Assessment	1-12	Structured log	Instrument to be adapted to any communication skill/task.

* Only speaking and listening materials are described.

OTHER	TECHNICAL INFORMATION RELIABILITY	VALIDITY	COMMENTS	AVAILABLE FROM
Content assists classroom teachers to plan and monitor oral language instruction. Some instruments would require training to use.	Unknown	Unknown	Title: Enhancing and Evaluating Oral Communication in the Primary, Intermediate and Secondary Grades, 1988	British Columbia Ministry of Ed., Victorial, B.C. Canada
	Unknown	Unknown	Title: Listening Profile & Listening Awareness Assessment Questionnaires	Calgary School Dist. Calgary, Alberta, Canada. Also, J. of International Listening Assoc., 2, 33-52, 1988.
Additional assistance with scoring would be required.	Fair-Good	Good	Title: Oral Communication, Assessment Program, (1981)	Glynn Co. School System, Brunswick, GA 31521. Also D. Rubin, U. of Georgia, Athens
	Unknown	Unknown	Titles: Competency-Based Measures for Grade 3 Performance Expectations (1987); Grade 10 CBM Technical Report (1988)	S. Chin-Chance, Hawaii State Dept. of Ed., 3430 Leahi Ave. Bldg. E., Honolulu, HI 96815
This is mostly a curriculum guide with ideas for assessment included	Unknown	Unknown	Titles: Speaking and Listening Activities in Illinois Schools, 1986; Write On Illinois, 1987	Illinois State Dept. of Ed., 100 N. First St. Springfield, IL 62777
Content of handbooks mainly to assist teachers to design instruction	Unknown	Unknown	Titles: A Guide to Developing Communication Across the Curriculum, 1989; A Guide to Curriculum Development in the Language Arts, 1986	Iowa Dept. of Ed. Grimes State Office Bldg., Des Moines, Iowa 50319

EDUCATIONAL AGENCIES*

AGENCY	DESCRIPTION			
	FOCUS	GRADES	FORMAT	CONTENT
Massachusetts Dept. of Ed.	Formal State Assessment	8 others?	Performance, Multiple-choice	2 instruments - 4 extended monologues scored analytically & a multiple-choice listening test.
Michigan Dept. of Ed.	Formal State Assessment	4,7,10	Multiple-choice	1 instrument covering critical listening using a variety of stimulus materials.
New Hampshire Dept. of Ed.	Formal State Assessment	5-12	Multiple-choice, short answer	2 levels covering listening comprehension using listening passages from real life.
New York Dept. of Ed.	Informal Classroom Assessment	K-12	Ratings, checklists, self- & peer evaluations	General Communication
	Formal State Assessment	12	Multiple-choice	Listening comprehension
New Zealand Council for Ed. Research	Formal Large Scale Assessment	3-12	Multiple-choice	Several levels covering listening comprehension using listening passages from real-life.
North Carolina Dept. of Ed.	Informal Classroom Assessment	1-2	Checklists	Several instruments covering speaking, oral language, listening and attitudes.
Ohio Dept. of Ed.	Informal Classroom Assessment	K-12	No sample instruments are provided.	

* Only speaking and listening materials are described.

OTHER	TECHNICAL INFORMATION RELIABILITY	VALIDITY	COMMENTS	AVAILABLE FROM
The report cited is not explicit enough to reproduce their training or the assessment.	Fair	Fair	Titles: State Speaking Assessment Instrument Technical Reports (1982, 1983); Massachusetts Test of Basic Skills: Listening (n.d.)	Massachusetts Dept. of Ed., Quincey Center Plaza, 1385 Hancock St. Quincey, MA 02169
	Unknown	Unknown	Title: Technical Report for the Objective Referenced Test for Critical Listening, 1980	Michigan Dept. of Ed., P.O. Box 420, Lansing, MI 48902
	Unknown	Unknown	Title: Listening Skills Assessment: Manual and Script, 1980	New Hampshire Dept. of Ed., Div. of Instruc., 64 N. Main St. Concord, NH 03301. Also in ERIC ED 236-657
	Unknown	Unknown	Titles: New York State English Language Arts Syllabus K-12, 1988,	New York State Ed. Dept., The Univ. of the State of NY, Albany, NY 12234
Part of a larger battery, that includes reading, writing, spelling and vocabulary	Unknown	Unknown	New York State Regents Comprehensive Exam in English. 1989.	
Part of longer battery that includes math and reading.	Fair-Good	Fair	Title: Progressive Achievement Tests, 1971. To be revised in 1993.	New Zealand Council for Ed. Research, P.O. Box 3237, Wellington, New Zealand (04)847-939
	Unknown	Unknown	Title: Communication Skills, Grades 1 and 2 Assessment, 1989.	North Carolina Dept. of Public Instruction, Raleigh, NC (919)733-3703
Handbooks cover recent research and sound instructional practices.	N/A	N/A	Title: Ohio English Language Arts Curr., 1985; Integrating Language Arts, 1985.	Ohio State Dept. of Ed., Div. of El & Sec. Ed., 655 Front St. Room 1005, Columbus, Ohio, 43266-0308

EDUCATIONAL AGENCIES*

AGENCY	FOCUS	GRADES	DESCRIPTION FORMAT	DESCRIPTION CONTENT
Ontario Ministry of Ed.	Informal Classroom Assessment	7-10	Multiple-choice, checklists, short answers, self-and peer-evaluations, performance.	Several instruments covering group discussions, extended monologues, listening comprehension, and use of mechanics.
Oregon Dept. of Ed.	Informal Classroom Assessment	K-12	Performance, multiple-choice, self- and peer-evaluations, checklists, short written responses.	Over 30 instruments that cover extended monologues, goup discussions, & listening comprehension.
Pennsylvania Dept. of Ed.	Informal Classroom Assessment	1-12	Performance, attitude survey	3 instruments covering extended monologues, classroom activities and student attitudes.
Saskatchewan Provincial Dept. of Ed.	Informal Classroom Assessment	3 others?		

* Only speaking and listening materials are described.

TECHNICAL INFORMATION				
OTHER	RELIABILITY	VALIDITY	COMMENTS	AVAILABLE FROM
	Unknown	Unknown	Title: The Ontario Assessment Instrument Pool: English II Intermediate Division, 1986.	Ontario Ministry of Education, Publication Centre, 880 Bay St. 5th floor, Toronto, Ontario, Canada M7A 1N8
Includes a lengthy procedure for assessing the process of putting together an oral presentation.	Unknown	Unknown	Titles: Integrated Assessment Model, 1988; Assessing Progress on the Common Curriculum Goals-Speaking and Listening, 1988; Procedures for Assessing Listening Skills, 1984.	Oregon State Dept. of Education, 700 Pringle Parkway, SE, Salem, OR 97310-0290
	Unknown	Unknown	Title: Speech in the Classroom: Assessment Instruments, 1980	Pennsylvania Dept. of Ed., PO Box 911, Harrisburg, PA 17126
The guides are not yet completed but will include both instructional and assessment ideas.	N/A	N/A	Title: Saskatchewan English Language Arts Curriculum, 1989.	Saskatchewan Ed., 2220 College Ave., Regina, Canada S4P 3V7

BIBLIOGRAPHY

BIBLIOGRAPHY

Arter, J.A., Deck, D.D. and Nickel, P. (1987). The Hawaii State Test of Essential Competencies, technical report. Northwest Regional Educational Laboratory, 101 SW Main, Suite 500, Portland, Oregon, 97204.

Backlund, P. (1985). Essential speaking and listening skills for elementary school students. *Communication Education, 34*, 185-195.

Backlund, P., Gurry, J., Brown, K. and Jandt, F. (1980). Evaluating speaking and listening skill assessment instruments: Which one is best for you? *Language Arts, 57*, 621-627.

Backlund, P.M., Brown, K.L., Gurry, J. and Jandt, F. (1982). Recommendations for assessing speaking and listening skills. *Communication Education, 31*, 9-17.

Barker, L.L. (1984). *Communication, 3rd Edition*. Englewood Cliffs, NJ: Prentice-Hall.

Bassett, R. E., Whittington, N. and Staton-Spicer, A. (1978). The basics in speaking and listening for high school graduates: What should be assessed? *Communication Education, 27*, 293-303.

Booth-Butterfield, M. (1986). Stifle or stimulate? The effects of communication task structure on apprehensive and non-apprehensive students. *Communication Education, 35*, 337-348.

Bostrom, R.M. and Waldhard, E.S. (1988). Memory models and the measurement of listening. *Communication Education, 37*, 1-13.

Buros, O.K. (1978). *The eighth mental measurements yearbook*. Highland Park, NJ: Gryphon Press.

Carbol, B.C. (1986). New approaches to language skills assessment. Paper presented at the Washington State Assessment Conference, Seattle.

Conoley, J.C. and Kramer, J.J. (Eds.) (1989). *The tenth mental measurements yearbook*. Lincoln, NB: University of Nebraska Press.

Conoley, J.C., Kramer, J.J. and Mitchell, J.V. (Eds.) (1988). *Supplement to the ninth mental measurements yearbook*. Lincoln, NB: University of Nebraska Press.

Devine, T.G. (1982). Listening skills schoolwide. ERIC ED 219 789.

Dickson, W.P. (Ed.) (1981). *Children's oral communication skills*. New York: Academic Press.

Dickson, W.P. (1981). Introduction: Toward a interdisciplinary conception of children's communication abilities. In Dickson, W.P. (Ed.) *Children's oral communication skills*, New York: Academic Press.

Fagan, W.T., Jensen, J.M. and Cooper, C.R. (1985). *Measures for research and evaluation in the English language arts, volume 2*. Urbana, IL: National Council of Teachers of English.

Faires, C.L. (1980). The development of listening tests. Paper presented at the Mid-South Educational Research Association Meeting, New Orleans, 1980. Also, ERIC ED 220 528.

Hammill, D.D., Brown, L. and Bryant, B.R. (1989). *A consumer's guide to tests in print*. Austin, TX: PRO-ED.

Hohl, S. and Cheney-Edwards, B. (1976). Listening comprehension grades 1-3. Cambridge, MA: Educator's Publishing Service.

Hutchinson, C., Prolitt A. and Munro, L. (1987). *User's guide for the English Language Skills Profile*. London: MacMillan Education.

Illinois State Board of Education (1987). Speaking and listening activities in Illinois schools -- Sample instructional and assessment materials. Illinois State Board of Education, Curriculum Improvement Section, 100 N. First Street, Springfield, Illinois 62777-0001.

Iowa Department of Education (1986). A guide to curriculum development in language arts. Iowa Department of Education, Grimes State Office Building, Des Moines, Iowa, 50319.

Iowa Department of Education (1989). A guide to developing communication across the curriculum. Iowa Department of Education, Grimes State Office Building, Des Moines, Iowa, 50319.

Jeroski, S., Diewert, J., Ford, C., MacLeod, C., Norris, N. and Robertson, L. (1988). Enhancing and evaluating oral communication in the primary grades: Teacher's resource package. British Columbia Department of Education, Victoria, B.C., Canada.

Joint Committee on Testing Practices (1988). Code of fair testing practices in education. American Psychological Association, 1200 17th Street NW, Washington D.C., 20036.

Karr, M. and Vogelsang, R.W. (1989). A comparison of the audio and pilot video versions of the Watson-Barker Listening Test: Forms A and B. Department of Speech Communication, Portland State University, P.O. Box 751, Portland, Oregon 97207.

Keyser, D.J. and Sweetland, R.C. (Eds.) (1985). *Test critiques*. Austin, TX: PRO-ED.

Larson, C.E. (1978). Problems in assessing functional communication. *Communication Education, 27*, 304-309.

Leary, M.R. (1988). Socially-based anxiety: A review of measures. In C.H. Tardy (Ed.), *A handbook for the study of human communication: Methods and instruments for observing, measuring, and assessing communication processes*, Norwood, NJ: Ablex.

Lederman, L.C. and Rubin, B.D. (1984). Systematic assessment of communication games and simulations: An applied framework. *Communication Education, 33*, 152-158.

Lundsteen, S.W. (1979). *Listening -- It's impact at all levels on reading and the other language arts*. National Council of Teachers of English, 1111 Kenyon Rd., Urbana, Illinois 61801. Also ERIC ED 169 537.

Massachusetts Department of Education (1982). Massachusetts State speaking assessment instrument. Massachusetts Department of Education, 1385 Hancock St., Quincy, MA 02169.

McCroskey, J.C. (1986). *An introduction to rhetorical communication*. Englewood Cliffs, NJ: Prentice-Hall.

McCroskey, J.C. and Daly, J.A. (Eds.) (1987). *Personality and interpersonal communication*. Newbury Park, CA: Sage.

Mead, N. A. (1978). Issues related to assessing listening ability. Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978. Also ERIC ED 155 759.

Mead, N.A. (1986). Listening and speaking skills assessment. In R.A. Berk (Ed.) *Performance Assessment Methods and Applications*, Baltimore: Johns Hopkins University Press.

Mitchell, J.V. (Ed.) (1985). *The ninth mental measurement yearbook*. Lincoln, NB: The University of Nebraska Press.

Ohio Department of Education (1985). Ohio English language curriculum. Ohio Department of Education, 65 South Front Street, Room 207, Columbus, Ohio, 43266-0308.

Phillips, G.M. (1983). A competent view of "competence." *Communication Education, 32*, 25-36.

Plattor, E. (1988). Assessing listening in elementary and junior high schools: A examination of four listening tests. *Journal of the International Listening Association, 2*, 33-52.

Powers, D. E. (1984). Considerations for developing measures of speaking and listening. College Board Report No. 84-5. ETS RR No. 84-18. College Entrance Examination Board, New York.

Reed, L. (1984). Assessing children's language skills. In C. Thaiss and C. Suhor (Eds.) *Speaking and writing K-12*, National Council of Teachers of English, 1111 Kenyon Rd., Urbana, Illinois 61801. Also ERIC ED 233 380.

Rubin, D.L. (1981). Using performance rating scales in large scale assessments of oral communication proficiency. In *Perspectives on the assessment of speaking and listening skills in the 1980's*, Northwest Regional Educational Laboratory, 101 SW Main, Suite 500, Portland, Oregon, 97204.

Rubin, D.L. and Rafoth, B.A. (1986). Oral language criteria for selecting listenable materials: An update for reading teachers and specialists. *Reading Psychology: An International Quarterly, 7*, 137-151.

Rubin, D. L. and Mead, N. A. (1984). *Large scale assessment of oral communication skills: Kindergarten through grade 12*. Speech Communication Association, 5105 Backlick Road, Suite E, Annandale, Virginia 22003. Also ERIC ED 245 293.

Rubin, R.B. (1982). Assessing speaking and listening competence at the college level: The Communication Competency Assessment Instrument. *Communication Education, 31*, 19-32.

Rubin, R.B. (1985). The validity of the CCAI. *Communication Monographs, 52*, 173-185.

Rubin, R.B. and Feezel, J. (1986). Elements of teacher communication competence. *Communication Education, 35*, 254-268.

Rubin, R.B. and Graham, E.E. (1986). Communication correlates of college success. *Communication Education, 37*, 15-21.

Rubin, R.B. and Roberts, C.V. (1987). Comparative examination and analysis of three listening tests. *Communication Education, 36*, 142-153.

Spandel, V. (1988). Listening Skills: Report 3. Assessing student progress on the common curriculum goals -- English language arts. Oregon Department of Education, 700 Pringle Parkway, Salem, Oregon. Also ERIC ED 298 519.

Spandel, V. (1988). Speaking skills: Report 2. Assessing student progress on the common curriculum goals -- English language arts. Oregon Department of Education, 700 Pringle Parkway, Salem, Oregon. Also ERIC ED 298 518.

Spandel, V. and Stiggins, R. (1989). Issues and trends in writing assessment. Northwest Regional Educational Laboratory, 101 S.W. Main, Suite 500, Portland, Oregon 97204.

Spitzberg, B.H. (1988). Communication competence; Measures of perceived effectiveness. In C.H. Tardy (Ed.), *A handbook for the study of human communication: Methods and instruments for observing, measuring, and assessing communication processes.* Norwood, NJ: Ablex.

Stiggins, R.J. (1981). Potential sources of bias in speaking and listening assessment. In *Perspectives on the assessment of speaking and listening skills in the 1980's*, Northwest Regional Educational Laboratory, 101 SW Main, Suite 500, Portland, Oregon, 97204.

Ting-Toomey, S. and Korzenny, F. (Eds.) (1989) *Language, communication and culture: Current directions.* Newbury Park, CA: Sage.

Tittle, D.K. (1989). Validity: Whose construction is it in the teaching and learning context? *Educational Measurement Issues and Practice, 8*, 5-13.

Utah State Board Of Education (1987). Language arts core curriculum, grades 7-12. Utah State Office of Education, 250 East 500 South, Salt Lake City, Utah, 84111.

Utah State Board Of Education (1987). Language arts core curriculum, grades K-6. Utah State Office of Education, 250 East 500 South, Salt Lake City, Utah, 84111

University of the State of New York (1988). English language arts syllabus K-12. State Education Department, Bureau of English and Reading Education, Bureau of Curriculum Development, Albany, New York 12238.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Kappan*, May, 1989, 703-713.

Wilkinson, A., Barnstey, G., Harma, P. and Swan, M. (1979). Assessing language development: The crediton project. ERIC ED 178 906.

Wolvin, A.D. and Coakley, C.G. (1985). *Listening*. Dubuque, IA: Wm. C. Brown, Publishers.

Wood, B.S. (1979). Development of functional communication competencies: Pre-K - Grade 6. Urbana, Illinois: ERIC.

GLOSSARY

GLOSSARY

Analytic Scoring A procedure for rating performances (writing samples, speaking, etc.) that uses a number of dimensions (such as content, organization, voice, sentence structure, and usage/mechanics).

Artificial Task An assessment task that has been developed specifically for the test and has aspects that are not typical of daily activities. Examples are: students have to give a three-minute speech on an assigned topic to a teacher who is rating their performance; students listen to various short passages without asking questions or taking notes, and answer questions about them; students role-play a job interview. It is the opposite of a naturalistic task, which is an activity engaged in by the student as part of ongoing life. There are different levels of artificiality depending on how closely the task mirrors real-life activities.

Audience For Communication That person or persons with whom one is interacting during a specific communication activity. It can be an individual, small group or large group; or peers, teachers, parents, employers, etc.

Cloze A type of test question in which a passage is presented to the students with some words deleted. The student supplies or chooses words that best complete the meaning of the selection.

Communication Anxiety Anxiety, apprehension or fear of communicating with others. This can be a general anxiety or can be focused toward specific situations or audiences.

Communication Competence The ability to communicate effectively for various purposes within various social contexts. This includes not only the knowledge of what words mean and how to construct messages, but also what constructions are most effective for various audiences, settings and purposes.

Construct Validity The degree to which an instrument measures an underlying psychological construct such as intelligence, motivation, or competence. If an instrument has construct validity, people's scores will vary as the theory underlying the construct would predict. For example, if an instrument has construct validity in the area of communication competence in speaking, then performance on the instrument would reflect performance in everyday situations.

Constructed Tasks See Artificial Tasks.

Content Of A Communication That which is being communicated about; for example, weather, cooking, health, school assignments, etc.

Content Validity How well the instrument samples from the skill domain of interest; how well student responses to the tasks in the instrument are a representative sample of all the possible tasks and responses in the curriculum area of interest.

Context For Communication The explicit or implicit setting, audience, purpose and content surrounding a communication. The context influences what will be effective.

Criterion Validity How performance on the instrument relates to other measures of the same thing; for example, other tests, grades, teacher ratings, etc.

Dichotomous Scoring A scoring procedure in which one indicates the presence or absence of a behavior or skill. An example would be a checklist of whether the student included various things in an extended monologue such as an introduction, major points, examples and a conclusion.

Discourse Mode The purpose for the communication, such as to convince (persuasive mode), explain (narrative mode), or tell (narrative mode).

Domain Of Skills The entire group of performances and abilities that constitutes a skill area, such as listening comprehension, persuasive speaking or writing.

Ecological Validity An instrument is ecological valid when it is used properly, results are perceived as being useful, and the use of the results does not promote negative side effects.

Extended Monologues A speech in which verbal interaction with the audience is not allowed. Speaking tasks developed for assessment purposes are usually extended monologues. This can add to the artificiality of many of these tasks. However, some real-life communication involves extended monologues, such as radio and TV reports.

Holistic Scoring A procedure for rating performances (writing, speaking, etc.) that uses a single score to indicate the overall quality of the piece.

Interactive Communication A communication activity in which people interact. This usually involves both speaking and listening. Examples are conversations, lectures in which students can ask questions and speeches in which audience feedback is allowed. This is the opposite of one-way communication in which messages are given or received, but no interaction is allowed (such as an extended monologue).

Linguistic Competence The sophistication of students with respect to the complexity of the language they can produce and understand. It would include such things as knowledge of vocabulary, the complexity of grammatical constructions used and understood, and the average length of student sentences.

Listenability The degree to which material contains certain features that are necessary if it is to be presented and understood verbally. Such features include simple sentence structure, a high degree of redundancy of information, thematic units that are resolved quickly, and a face-to-face style of language.

Objective Format Tests Any assessment format in which students choose answers rather than produce answers. Examples are multiple-choice, matching and true-false.

One-Way Communication A communication activity in which people do not interact. Examples are extended monologues and listening in which one is not able to interact with the presenter (e.g., radio and TV reports).

Performance Format Tests Any assessment format in which the responses are produced as opposed to being chosen. Examples are short answers, performance of tasks, telling a story and summarizing what is heard.

Primary Trait Scoring A procedure for rating performances (writing, speaking, etc.) which results in a single score that indicates the overall effectiveness of the writing for the purpose intended. For example, a persuasive speech would be rated on how well it persuades, while an expository speech would be rated on how well it explains something.

Purposes For Communication The reasons why the communication is taking place. These could include social necessity, obtaining information, recreation, or persuasion.

Reliability The degree of consistency between two measures of the same thing. Test-retest reliability is the degree to which measurements are consistent across time. Internal consistency reliability is the degree to which the items on the test tend to measure the same thing. Interrater reliability is the extent to which ratings from different persons are the same.

Setting For A Communication One component of the context for a communication. The setting includes group size, the formality of the occasion, interactive or one-way communication and amount of preparation (impromptu or not).

Skills In Concert An assessment situation in which students must use a variety of skills in concert in order to achieve a goal. An example is delivering a speech. Students have to not only present the material desired but also be responsive to the audience.

Skills in Isolation. An assessment situation which singles out skills for separate measurement. An example is a listening comprehension test in which one item measures main idea and the next covers mood.

Validity The extent to which an instrument measures what it claims and can be used for the purposes stated. Content, criterion and construct validity all contribute to the overall judgment of validity.

INDEX

INDEX

General Instruments — Page

Class Apprehension About Participation Scale 54
College Outcome Measures Program 36
Communication Competency Assessment Instrument 33

Diagnosis of Group Membership 68

English Language Skills Profile 44
Evaluating Communicative Competence 91

Fullerton Language Test for Adolescents 85

Hunter-Grundin Literacy Profiles 69

Interactional Competency Checklist 54
Interpersonal Language Skills Assessment 91

Jones-Mohr Listening Test 70

Kentucky Comprehensive Listening Test 91

Language Communication Skills Task 55
Language Inventory for Teachers 86
Language Proficiency Test 70

Notebook Communication Game 56

Personal Report of Communication Apprehension 56
Profile of Nonverbal Sensitivity 47

Repairs of Misunderstandings During Communication 72

Test of Adolescent Language-2 87
Test of Early Language Development-2 88
Test of Implied Meanings 73
Test of Language Development-2 89
Test of Pragmatic Skills 91
Two Referential Communication Tasks 57

Utah Test of Language Development-3 90

Watson-Barker High School Listening Test 49
Willingness to Communicate Scale 57

Achievement Tests **Page**

California Achievement Test (CAT) 60
CIRCUS 30
Comprehensive Test of Basic Skills (CTBS) 60
Comprehensive Test Program (CPT II) 61

Diagnostic Achievement Battery 40

Iowa Test of Basic Skills (ITBS) 62

Language Diagnostics Test 62
Listening Test (CAT & CTBS) 63

Metropolitan Achievement Test (MAT-6) 63
Metropolitan Readiness Test (MRT) 64

National Achievement Test (NAT) 64
National Test of Basic Skills (NTBS) 65

Stanford Achievement Test (SAT) 65
Stanford Early School Achievement Test (SESAT) 66
Survey of Basic Skills 66

Tests of Achievement and Proficiency 66

Anthologies of Informal Tools **Page**

Assessing Children's Speaking, Listening and Writing Skills 68

Evaluating Classroom Speaking 69

Listening Comprehension, Grades 1-3 71

Listening: Its Impact At All Levels on Reading and the Other Language Arts 72

Listening Skills Schoolwide 72

Speaking Skills: Report 3. Assessing Student Progress on the Common Curriculum Goals 73

Educational Agencies	Page
British Columbia Ministry of Education	75
Calgary School District	75
Glynn County, Georgia	76
Hawaii State Department of Education	77
Illinois State Department of Education	77
Iowa State Department of Education	78
Massachusetts State Department of Education	78
Michigan State Department of Education	79
New Hampshire State Department of Education	80
New York State Department of Education	80
New Zealand Council for Ed. Research	81
North Carolina State Department of Education	82
Ohio State Department of Education	82
Ontario Ministry of Education	83
Oregon State Department of Education	83
Pennsylvania State Department of Education	84
Saskatchewan Provincial Department of Education	84